

MB

Journal of
**Geophysical
Research**

VOLUME 65

FEBRUARY 1960

NUMBER 2

**THE SCIENTIFIC PUBLICATION
OF THE AMERICAN GEOPHYSICAL UNION**

Journal of Geophysical Research

An International Scientific Publication

OFFICERS OF THE UNION

LLOYD V. BERKNER, *President*
F. W. REICHELDERFER, *Vice President*
A. NELSON SAYRE, *General Secretary*
WALDO E. SMITH, *Executive Secretary*

OFFICERS OF THE SECTIONS

Geodesy

CHARLES PIERCE, *President*
FLOYD W. HOUGH, *Vice President*
BUFORD K. MEADE, *Secretary*

Seismology

LEONARD M. MURPHY, *President*
JAMES A. PEOPLES, JR., *Vice President*
BENJAMIN F. HOWELL, JR., *Secretary*

Meteorology

THOMAS F. MALONE, *President*
GORDON E. DUNN, *Vice President*
WOODROW C. JACOBS, *Secretary*

Geomagnetism and Aeronomy

L. R. ALLDREDGE, *President*
C. T. ELVEY, *Vice President*
J. HUGH NELSON, *Secretary*

Oceanography

WALTER H. MUNK, *President*
DONALD W. PRITCHARD, *Vice President*
EUGENE C. LAFOND, *Secretary*

Volcanology, Geochemistry, and Petrology

ALFRED O. C. NIER, *President*
FRANCIS J. TURNER, *Vice President*
IRVING FRIEDMAN, *Secretary*

Hydrology

WALTER B. LANGBEIN, *President*
WILLIAM C. ACKERMANN, *Vice President*
CHARLES C. McDONALD, *Secretary*

Tectonophysics

PATRICK M. HURLEY, *President*
LOUIS B. SLICHTER, *Vice President*
H. RICHARD GAULT, *Secretary*

BOARD OF EDITORS

Editors: PHILIP H. ABELSON and J. A. PEOPLES,

ASSOCIATE EDITORS

1960

HENRY G. BOOKER	WALTER B. LANGBEIN
E. C. BULLARD	ERWIN SCHMID
JULE CHARNEY	HENRY STOMMEL
GEORGE T. FAUST	J. TH. THIJSSSE
DAVID G. KNAPP	A. H. WAYNICK

J. TUZO WILSON

1960-1961

HENRI BADER	T. NAGATA
K. E. BULLEN	FRANK PRESS
CONRAD P. MOOK	A. NELSON SAYRE
WALTER H. MUNK	MERLE A. TUVE

JAMES A. VAN ALLEN

1960-1962

JULIUS BARTELS	TOR J. NORDENSON
E. G. BOWEN	E. N. PARKER
JOHN E. CHAPPELEAR	GEORGE P. RIGSBY
G. D. GARLAND	WALTER O. ROBERTS
GORDON J. F. MACDONALD	C. N. TOUART
L. A. MANNING	JAMES R. WAIT

This Journal welcomes original scientific contributions on the physics of the earth and its environment.

Manuscripts should be transmitted to J. A. Peoples, Jr., Geology Department, University of Kansas, Lawrence, Kansas. Authors' institutions, if in the United States or Canada, are requested to pay a publication charge of \$25 per page, which, if honored, entitles them to 100 free reprints.

Subscriptions to the *Journal of Geophysical Research* and *Transactions, AGU* are included in membership dues.

Non-member subscriptions, *Journal of Geophysical Research*..\$30 for back Volume of 1959, \$5 for this issue; \$20 for the calendar year 1960.

Non-member subscriptions, *Transactions, AGU*.....\$4 per calendar year, \$1.25 per copy.

Subscriptions, renewals, and orders for back numbers should be addressed to American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington, D. C. Suggestions to authors are available on request.

Advertising Representative: Howland and Howland Inc., 230 Park Ave., New York 17, N. Y.

Beginning with the January 1959 issue (Vol. 64, No. 1) the *Journal of Geophysical Research* is published monthly by the American Geophysical Union, the U. S. National Committee of the International Union of Geodesy and Geophysics organized under the National Academy of Sciences-National Research Council as the U. S. national adhering body. Publication of this journal is supported by the National Science Foundation and the Carnegie Institution of Washington. The new monthly combines the type of scientific material formerly published in the bi-monthly *Transactions, American Geophysical Union*, and the quarterly *Journal of Geophysical Research*. The *Transactions, American Geophysical Union* will continue as a quarterly publication for Union business and items of interest to members of the Union.

Published monthly by the American Geophysical Union from 1407 Sherwood Avenue, Richmond, Virginia. Second class postage paid at Richmond, Virginia.

micrometeorological

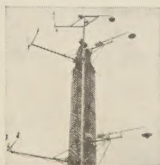
Micrometeorological tower with three stations of wind instruments rises 150 ft above the National Aeronautics and Space Administration's Plumbrook reactor facility at Sandusky, Ohio. This is an example of Beckman & Whitley systems engineering. Standard components are integrated into an installation designed to meet a specialized set of requirements.

systems engineering

Systems engineering personnel from San Carlos provided sensor groups capable of feeding temperature-gradient, ambient-temperature, wind-speed, and wind-direction data to a flexible array of indicating and recording instruments in two special equipment racks.

To provide convenience in access to the sensor equipment, the system was installed and field tested by this group of specialists on a retractable telescoping tower which can be lowered manually or electrically to 32 ft, as shown below.

Your meteorological-system problems can be solved in an efficient, effective way by this group. Their experience is available to you.



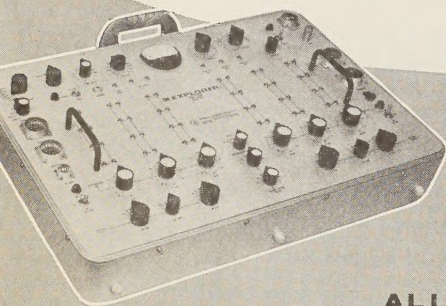
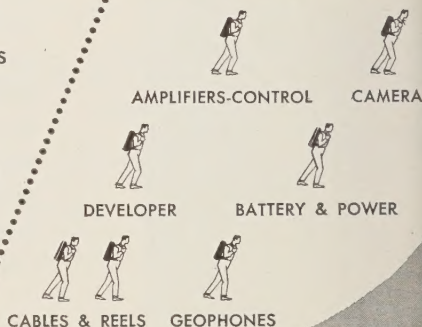
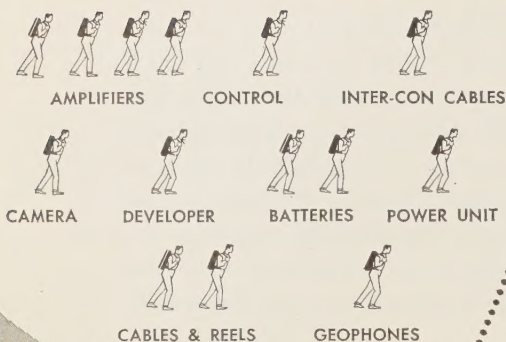
by

Beckman & Whitley

SAN CARLOS 15 • CALIFORNIA

TYPICAL FULL RANGE 24-trace seismograph

EXPLORER FULL RANGE 24-trace seismograph



**Cut manpower in half
...double production with**

EXPLORER

ALL-TRANSISTORIZED SEISMOGRAPH

In some of the world's toughest assignments, the EXPLORER Model 8000 24-trace All-Transistorized Seismograph has effected reductions in recording personnel of 50 percent from the number required for the typical full range portable systems. Due to the reduced weight and number of packages for amplifiers, batteries and interconnecting cables, only seven men are required for the 8000 system. In foreign operations, the additional savings in manpower will be several times the number of packages eliminated due to release of support personnel.

Of most importance, it is reported that

the decreased time involved at and between each recording set-up with the EXPLORER Seismograph has approximately doubled production, and, of course, cut the cost per profile in half.

Furthermore, the 8000 system is effective over a wide range of frequencies, from high resolution to refraction. You have a wide selection of AGC fixed and TVG gain controls.

The EXPLORER is the only completely *field-proved* All-Transistorized Seismograph System. Let the TI engineer show you how it pays for itself in *reduced operating costs, increased production, and unequalled reliability.*

***Now operating in Sumatra, Peru, Columbia,
Mexico, France, Canada and South Louisiana.**

Write for Bulletin No. S-324



TEXAS INSTRUMENTS INCORPORATED

GEOSCIENCES & INSTRUMENTATION DIVISION
3609 BUFFALO SPEEDWAY • HOUSTON 6, TEXAS • CABLE: TEXINS



Lockheed Missiles and Space Division Offers

PHYSICISTS

Unusual Opportunities in Research and Development

■ Important career positions are available at Lockheed Missiles and Space Division at its new facilities on the beautiful San Francisco Peninsula — one of the choicest living areas in the nation.

Headquarters for the Division are at Sunnyvale, California, with Research and Development facilities located in the Stanford Industrial Park in nearby Palo Alto. Equipment and facilities are completely modern and include one of the largest computer centers in the world.

Interesting opportunities exist for physicists to contribute to the solution of new problems in the fields listed at the right. Please write:

Research and Development Staff, Dept. B-59,
962 West El Camino Real, Sunnyvale, California.
U.S. citizenship or existing Department of
Defense clearance required.

SPACE PHYSICS

Advanced degree preferred, for work in basic research on the physics of the earth's upper atmosphere and beyond. Typical projects include: measurement of atmospheric composition and density at satellite altitudes; laboratory experiments on upper atmospheric atomic and molecular reactions; hydromagnetic interactions with the earth's magnetic fields; simulation and study of meteor impacts; and particle radiation.

INFRARED AND OPTICS

Advanced degree in E.E. or physics preferred, with experience in infrared systems research and development, or electrical engineers with background in electronics information theory; servomechanisms; specialized circuitry, as in low-level voltage circuits, or physicists with background in optics or semiconductors. For work in infrared physics research; advanced systems development, or physical measurements in infrared. Research is also being conducted in optical devices and systems, including scanners, encoders, detectors, and read-out devices.

SOLID STATE DEVICES

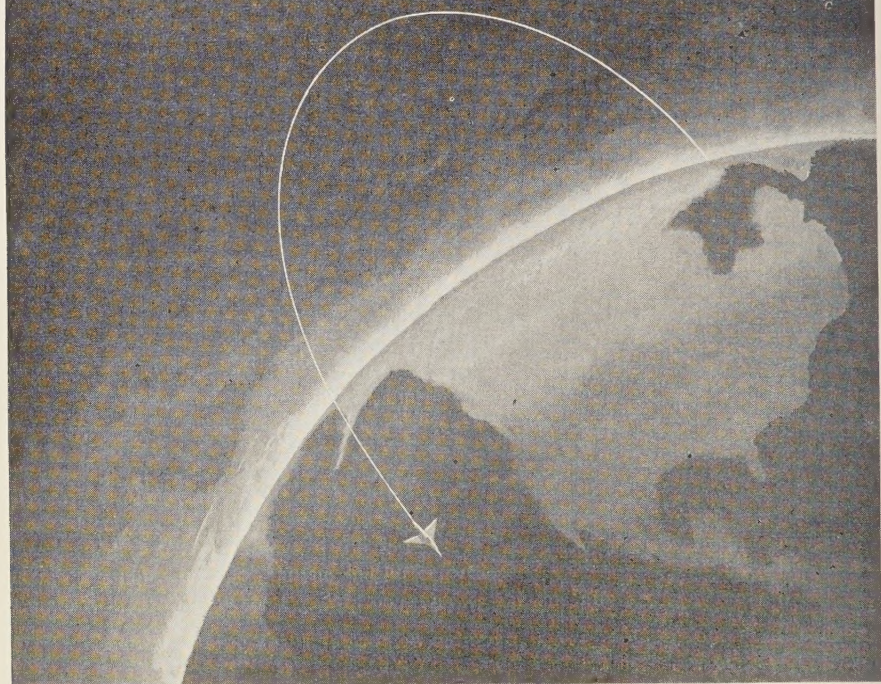
Advanced degree required, Ph.D. preferred, in E.E. or physics and evidence of creative, original work through published articles, patents or superior Ph.D. theses for research work in one or more of the following: thermoelectric; photovoltaic; lumistor; ferrite; logic components; sensor; thermistor; or cryogenic devices. Also, materials analysis and evaluation; processing techniques design and development of novel electronic devices and components; circuit analysis; circuit topology; or microminiaturization.

Lockheed / **MISSILES AND SPACE DIVISION**

Systems Manager for the Navy POLARIS FBM; the Air Force AGENA Satellite in the DISCOVERER Program and the MIDAS and SAMOS Satellites; Air Force X-7; and Army KINGFISHER

SUNNYVALE, PALO ALTO, VAN NUYS, SANTA CRUZ, SANTA MARIA, CALIF. • CAPE CANAVERAL, FLA. • ALAMOGORDO, N. M. • HAWAII

We can now measure gravity from the air



Another achievement in gravity measurement from LaCoste & Romberg

Now for the first time, gravity surveys of inaccessible areas can be made from the air, using a LaCoste & Romberg airborne gravity meter. This new meter requires no gyroscopic stabilization platform. Its accuracy is better than 10 milligals.

Commercial airborne surveys are now available from Fairchild LaCoste Gravity Surveys, Inc.

LaCoste & Romberg 6606 NO. LAMAR AUSTIN, TEXAS

Manufacturers of airborne, submarine, shipborne and surface gravity meters for both exploration and geodetic surveys

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

1. SUBSURFACE MAPPING

PREPARATION AND INTERPRETATION

by MARGARET S. BISHOP, *University of Houston*. Vital to the oil geologist, this book stresses the new stratigraphic-type maps that are so valuable in explorations for oil. It also explains straight-to-the-point the various other kinds of maps that may be made to present subsurface data, as well as a clear discussion on how to interpret these maps. This is not a textbook; it is rather an immediately practical volume for the technician and engineer. 1960. 198 pages. \$5.75.

2. LITHOFACIES MAPS

An Atlas of the United States and Southern Canada

by L. L. SLOSS, E. C. DAPPLES, and V. C. KRUMBEIN, *all at Northwestern University*. Shows by contour and two-color patterns the thickness of strata, their lithologic aspects, and their surface and subsurface distribution. Includes about 150 rock- and time-stratigraphic units in the U. S. and Southern Canada. 1958. Ready Spring 1960.

TRIAL ORDER

JOHN WILEY & SONS, Inc.

605 PARK AVE. SOUTH, NEW YORK 16, N. Y.

2

JGR-260

Send on 10 days' approval the books circled. Within 10 days of receipt I'll remit full price or return books postpaid.

Name.....

Street.....

City..... Zone..... State.....

Check here to save postage. Enclose full amount with order and we pay postage. Same return privilege.

GEOTECH GROUND-BASED ANALOG DATA—TRANSMISSION EQUIPMENT

- ▶ **CAPACITY**—1 to 7 channels of low-frequency analog data.
- ▶ **TRANSMISSION CIRCUIT**—a single voice-frequency telephone, radio, or microwave circuit.
- ▶ **TRANSMISSION RANGE**—coast to coast if required.
- ▶ **DEPENDABILITY**—proven, continuous year-round operation.
- ▶ **ECONOMY**—building-block components; ± 100 volt output of discriminator eliminates need of DC amplifier; can use a leased commercial circuit.
- ▶ **COMPONENTS**—FM telemetering multiplexer: output $z=600$ ohms, $8\frac{3}{4}"$ h x $19"$ w x $15"$ d; voltage—controlled oscillators: RDB/IRIG subcarrier channels, $\pm 2\frac{1}{2}$ volts input produces $\pm 7\frac{1}{2}\%$ deviation, $4\frac{7}{8}"$ h x $3\frac{7}{8}"$ w x $1\frac{5}{8}"$ d; discriminators: input z —over 1 megohm, output $z=0$ ohms, $5\frac{1}{4}"$ h x $19"$ w x $16\frac{3}{8}"$ d.



RUGGED SENSITIVE GALVANOMETERS SERIES 2980



- ▶ **PERIODS**—1 to 90 seconds.
- ▶ **SENSITIVITY**— 9×10^{-11} amps/mm/meter typical at 90 seconds.
- ▶ **STABILITY**—new fast-stabilizing design provides extremely small drift.
- ▶ **ADJUSTABLE**—CDRX, sensitivity, leveling, and horizontal light spot position.

SERIES 4100

- ▶ **FREQUENCIES**—1 to 50 cps.
- ▶ **SENSITIVITY**—up to 3×10^{-9} amps/mm/meter.
- ▶ **SUSPENSION**—separate, insulated suspension frame rotates, does not disturb ribbon.
- ▶ **ADJUSTABLE**—air gap, natural frequency, horizontal and vertical light spot position.

FOR INFORMATION WRITE:



THE GEOTECHNICAL CORP.

3401 Shiloh Road • Garland, Texas

P. O. Box 28277 • Dallas 28, Texas

**FOUR IMPORTANT NEW BOOKS
FROM MCGRAW-HILL**

IGNEOUS AND METAMORPHIC PETROLOGY, *New Second Edition*

By FRANCIS J. TURNER and JOHN VERHOOGEN, *both of the University of California, Berkeley*. 694 pages, \$12.00

As before, the book represents a unified general impression of origin and evolution of rocks that have crystallized, or have been profoundly modified, at high temperatures. It is correlated with modern conceptions as to the nature and prevailing physical conditions of the earth's crust and of the outer part of the underlying mantle. Igneous and metamorphic phenomena have been treated, in a single volume, as partially dependent on each other, and as being controlled by the same general physico-chemical principles.

INTRODUCTION TO GEOPHYSICAL PROSPECTING, *New Second Edition*

By MILTON B. DOBRIN, *Triad Oil Company, Ltd. Calgary, Alberta, Canada*. 446 pages, \$9.50.

A thorough revision of a highly successful text. It is designed to present the principles of current techniques of geophysical prospecting for oil and minerals to students and technical personnel employed in the fields of petroleum and mineral exploration. The book covers all the major methods of geophysical prospecting. For each method it discusses fundamental physical principles, instruments, field techniques, reduction of data, interpretation, and examples showing results of actual surveys.

GEOLOGY: Principles and Processes, *New Fifth Edition*

By WILLIAM H. EMMONS, IRA S. ALLISON, *Oregon State College*; GEORGE A. THIEL, *University of Minnesota*; and CLINTON R. STAUFFER, *California Institute of Technology*. Ready in April.

A textbook for a beginning one-semester course in Physical Geology. It has been a standard and leading text for this field for over 27 years. This new fifth edition has been the most extensive of all its revisions. It is receiving special editing treatment and the result will be a completely rewritten text, with entirely new artwork done by a professional scientific illustrator, and a skilled use of two colors throughout. The text is suitable for both terminal cultural courses, or for introductory courses for majors in the subject.

PRINCIPLES OF PETROLEUM GEOLOGY, *New Second Edition*

By WILLIAM L. RUSSELL, *Agricultural and Mechanical College of Texas*. Ready in March.

This new edition of one of the most important books in this area has been carefully revised; much new material has been added, and all other information has been brought fully up to date. The first 17 chapters deal with principles; the remaining 12 chapters cover the specialized methods or techniques used by petroleum geologists. The book will be of value, therefore, not only to students, but also to those involved in the oil industry.

Send for Copies on Approval

McGRAW-HILL BOOK COMPANY, INC.

330 West 42nd Street

New York 36, N. Y.

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers

SPRENGNETH'S DIRECT WRITING VISUAL ORDER PROVIDES 24 HOUR REGISTRATION.

For a moderate initial investment, seismological laboratories can obtain this superior drum-type recorder. Its advantages include continuous registration and easy visual access to all information on the recorder. It also requires less storage space for records than a pen recorder.

Drum is completely enclosed to protect against dust or accidental damage. For greatest possible convenience in changing records, the large, curved plastic cover can be fully opened.

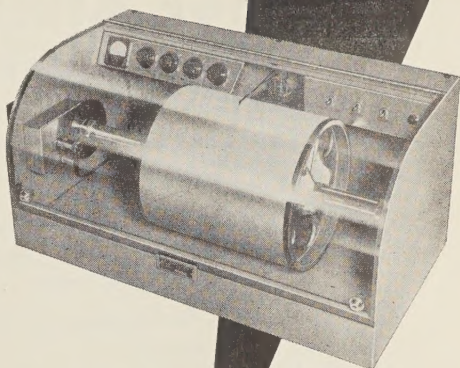
The pen drive galvanometer is a high torque, frictionless, torsion-type moving coil system. It operates in the field of large Alnico V magnet with special pole pieces and core.

Pen-drive is rigidly mounted on inside of back panel of recorder. Ink-well is mounted on the axis of moving coil to eliminate inking problems.

Recorder box is of heavy steel and has a beautiful, baked enamel finish.

SPECIFICATIONS:

Dimensions: 31" long, 16" wide, 15" high (not including drum housing) — Weight: 90 lbs. net — Drum Speed: 60 or 120 mm/min. — Translation Rate: 2.5 or 5 mm/rev. — Power Requirements: 110 V. or 240 V. AC, 50 or 60 cycles. — Paper Size: 36" x 12" — Pen Galvanometer Sensitivity: 100 milliamperes/mm.



Series VR-40-0

Shown with controls as used with our VR-30-A Amplifier.

Write for complete technical information today.

Internationally Known Mfrs. of Seismological, Geophysical Instruments.
W. F. SPRENGNETH INSTRUMENT COMPANY INC.
4567 SWAN AVE. ST. LOUIS 10, MO.

8
7
6
5

N 14.008 e 19	
C 12.011 e 0032	
B 10.822 e 755	B8 0.5 s B* 14 (2 e 3) E 18
Be	Be7 5.3 e 4

ISOTOPES, INC.

123 Woodland Avenue, Westwood, New Jersey

MEASUREMENT OF TRITIUM

in natural waters

This New Service Enables

- **METEOROLOGISTS**
to study mixing, transport and diffusion of water vapor in the atmosphere.
- **OCEANOGRAPHERS**
to study rates of movement and mixing of surface ocean water.
- **HYDROLOGISTS**
to study surface runoff, mixing in lakes and reservoirs, infiltration rates and ground water circulation.

Write for further information.

Other services include low level C^{14} measurements and radium determinations by the emanation method.

Physics of the Upper Atmosphere

Edited by J. A. RATCLIFFE, Cavendish Laboratory, Cambridge

February 1960, about 600 pp., illus., \$14.50

The Thermosphere—The Earth's Outermost Atmosphere

By SYDNEY CHAPMAN

The Properties and Constitution of the Upper Atmosphere

By M. NICOLET

The Upper Atmosphere Studied by Rockets and Satellites

By HOMER E. NEWELL, JR.

The Sun's Ionizing Radiations

By HERBERT FRIEDMAN

The Airglow

By D. R. BATES

General Character of Auroras

By D. R. BATES

The Auroral Spectrum and Its Interpretation
By D. R. BATES

Radar Studies of the Aurora

By HENRY G. BOOKER

The Ionosphere

By J. A. RATCLIFFE and K. WEEKES

The Upper Atmosphere and Geomagnetism

By E. H. VESTINE

The Upper Atmosphere and Meteors

By J. S. GREENHOW
and A. C. B. LOVELL

Advances during the I.G.Y. 1957/58

By the CONTRIBUTORS

AUTHOR INDEX-SUBJECT INDEX.

ACADEMIC PRESS, New York and London

111 FIFTH AVENUE, NEW YORK 3, NEW YORK

17 OLD QUEEN STREET, LONDON, S.W. 1



BULLETIN (IZVESTIYA), ACADEMY OF SCIENCES, U.S.S.R.

GEOFYSICS SERIES

Subscriptions for 1958 volume now available

This monthly Russian publication, perhaps the leading journal of Geophysics of the U.S.S.R., is being translated and published in an English edition for the year 1958 by the American Geophysical Union. The twelve numbers in Russian cover 1536 pages. Published with the aid of a grant from the National Science Foundation.

Send subscriptions now to

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Avenue, N.W.

Washington 5, D. C., U. S. A.

Subscription rates: \$25.00 for the volume of 12 numbers (\$12.50 for individuals subscribing for personal use; introductory offer)

Numbers will be mailed as issued.

The English edition of this publication for 1957 has been translated and published for the American Geophysical Union by Pergamon Press. This volume may also be ordered through the American Geophysical Union at a price of \$25.00 plus a service charge of \$3.00. The March 1959 issue of the *Transactions*, AGU, carries the titles of the papers of the first nine numbers of this volume, and the June 1959 issue carries the titles of the papers of the last three issues. It is anticipated that subsequent issues will carry the titles in the 1958 volume.

Journal of GEOPHYSICAL RESEARCH

VOLUME 65

FEBRUARY 1960

No. 2

INTERNATIONAL SYMPOSIUM ON ELECTRONIC DISTANCE-MEASURING TECHNIQUES¹

Introductory Remarks

ANTONIO MARUSSI

University of Trieste, Italy

Mr. President, Ladies, and Gentlemen: Taking the chair of this symposium on the techniques for direct measurement of distance in geodesy, tradition requires that I should do two things—the first one very pleasant and no mere formality, namely to extend to you my greetings, and those of the whole of Section I of the International Association of Geodesy, under whose auspices this symposium is being held.

My second task is to say a few words by way of introduction, and I have to confess that I could find this a great deal less pleasant, because I have no skill in the art of talking for the king's sake, were it not for the fact that this to is no mere formality. Fortunately for me I

have something to say which comes from the bottom of my heart, and which expresses my most profound convictions.

First of all, then, a very warm greeting to all of you, and every good wish for the work ahead. Above all, greetings and heartfelt thanks to our American hosts who have welcomed us to their incomparable capital.

Washington is dear to us all, precisely because of its spacious vistas that so exactly reflect the ideal concepts upon which this young nation is based and which have made her great. The realization of these concepts in the architecture of Washington is surely symbolic of the way they have found expression in so many aspects of American culture and life.

¹The Central Bureau of the International Association of Geodesy has recognized that there is not adequate time during a General Assembly of the U. G. G. to hold symposia on various scientific problems in geodesy. Therefore, a series of symposia has been planned for the three-year period between the General Assemblies. The American Geophysical Union, through the Section of Geodesy, invited the I. A. G. to hold a symposium on electronic distance-measuring techniques in Washington, D. C., May 5-12, 1959. The meetings were held in the Auditorium of the Department of Commerce. The proceedings of the symposium were collected and edited by C. A. Whitten and Erwin Schmid of the U. S. Coast and Geodetic Survey, Washington, D. C.

It is not feasible to publish the complete discussion. Some of the papers have been condensed or abstracted because of space limitations and

prior publication of part of the material. A paper on *TransAtlantic Trilateration Nets* prepared by Maurice Ewing, J. L. Worzel, and Manik Talwani was presented by Worzel. This paper was published in 1959 under the title *Some Aspects of Physical Geodesy* in *Geophysical Monograph No. 4* of the American Geophysical Union. The material presented by Max Kneissl of Munich, Germany, has been published under the title *Normalstrecke, Basis und Basisvergrößerungsnetz, München-Ebensberg Anlage und Vermessungsergebnisse 1958*, by Bayerische Akademie der Wissenschaftlichen, *Mathematisch-Naturwissenschaftliche Klasse, Abhandlungen Neue Folge, Heft 97*. In other instances speakers did not present material for publication. If readers wish to have additional information on any subject that has been abstracted, it is suggested that the request for it be submitted to the particular author.

We could certainly never have wished for a more appropriate meeting place for a symposium that aims to look so boldly and eagerly into the future.

A special expression of gratitude goes to our friend and colleague, the Vice-President of the International Association of Geodesy, Mr. Charles Whitten, who, together with his brilliant associates, has taken on the burden of organizing and directing the course of this meeting. I also wish to express our heartiest appreciation to the American scientific committees, the international organizations, and the business firms that have given moral and material support to this project. Finally, our thanks go to the American Geophysical Union, which has included the symposium in the program of its annual meeting.

Turning to the subject of the symposium itself, there is certainly no need for me to waste words by commenting on its importance, for that is obvious to us all. Nor do I propose to outline the things that you will see and hear in detail during its course. I should like, rather, to consider these new techniques of direct measurement in space, which today are to be presented publicly for the first time with an abundance of material and a wealth of detail, and I should like to try to fit them into the framework of the historical development of geodesy.

For countless generations, in antiquity, man considered the earth to be a flat, two-dimensional surface. The third dimension just barely came into the picture because of the elevations in the earth's surface, but obviously these were small in comparison with the other two dimensions considered.

Outside these two, there seemed in fact to be only a mythological world; below, a province that belonged to the study of demonology rather than to logical speculation, while above there was another world of fantasy, the seat of divinity and of supernatural phenomena, as inaccessible as the underworld to scientific investigation.

Even when it was realized that the earth was a sphere, the step forward was by no means so great as it might seem. The idea of two dimensions, length and breadth, gave way to the concept of longitude and latitude—but these

continued to dominate the scene, to the virtual exclusion of any idea of the third dimension.

When did man finally grasp the idea of the third dimension?—an idea that nowadays is much part and parcel of our intuitions that take it utterly for granted, as if we were born with the concept, instead of having had to acquire it as the final fruit of long experience.

The case of the third dimension is analogous with that of many other discoveries—just as Neptune was first discovered with the aid of Leverrier's calculations and only later picked out in the skies by Galle's telescope; just as electromagnetic waves were first discovered mathematically on paper by Maxwell, and then confirmed in reality by Hertz; just as many chemical elements were predicted together with their properties, thanks to Mendelyeev's law, before analysis revealed them in the laboratory—so, in the same way, we can say that in the conquest of the third dimension logical speculation and the intuition of a few supreme geniuses counted for more than the sum total of experience of millions of men who had lived before them. The names of these masterminds are Galileo, Kepler, and Newton, who with their entirely abstract speculations, supported by the barest amount of experience, radically altered our concept of the world in which human beings move, and who gave the earth its due place in the system of the universe. In this way the third physical dimension was added to the patrimony of future generations.

The geodetic sciences were not slow to avail themselves of this patrimony. It was thus possible to establish the physical principles that govern the appearance of our planet, and from these it was possible to construct coherent theories about the geometric structure of the field of forces which the planet generates and which, in their turn, condition its shape and the distribution of its masses.

With these conquests in the realm of theoretical progress in the field of practical geodesy could not manage to keep pace. It was still rooted in methods leading to the consideration of two dimensions only. Even nowadays, for that matter, you still hear it said quite often that geodesy is the science that occupies itself with the study of the shape and size of the earth when in reality it would be much more a

appropriate to say that it is concerned with the picture (whether geometric or dynamic) of the field of forces generated by the earth in space.

All practical geodesy from the end of the nineteenth century to the present day has ignored the third dimension, and the attempt has been made to limit its field of research to that of a particular surface—an ellipsoid, when schematic representation is required or when approximation will do for practical purposes—otherwise a geoid, when experimental research goes beyond these requirements.

And if the third dimension happens to crop up, then the first reaction in practical geodesy is to get rid of it as quickly as possible, by the simple process of bulldozing it away with the aid of a mass of corrections and reductions which every treatise of geodesy is full and over which, even today, violent arguments still arise, though the same corrections are used in practice without the slightest compunction.

It might be objected that what I have said does not correspond exactly with the truth, since geodesy has been concerned for centuries with surrounding space, by way of that particular branch of study called geodetic astronomy. I quite agree, but then I must point out that in this case the space surrounding us is not considered as the realm of 'objects,' but rather 'directions' toward points considered infinitely distant, such as the stars. And the directions of space still represent a variety of only two dimensions, which can in fact be identified with the surface of a sphere—in other words, the vault of the stars.

Now, however, things are at last beginning to change. The third dimension which was concentered in the mind in the age of Galileo and Newton is now being conquered materially in our own age—in fact in the past few years of our lives.

The space which surrounds us and which till now, as I have said, has been the realm of abstract speculations, or at the most of 'directions,' is now being penetrated and explored by objects put there by man.

Another object, already to be found there—the moon—until only a short while ago was as remote from us as any of the other bodies in space, but now has been brought close to us,

thanks to the ever-growing power of our means of observation. These now make it possible for us to regard the moon as a part of terrestrial space. In the same way, the observation of artificial satellites, of eclipses of the sun and occultations of the stars by the moon, have become procedures capable of rendering geodetic information of the greatest interest.

Fortunately for us the artificial satellites and the moon fly a bit too high for anyone to get the bright idea of bulldozing them with reductions to the geoid, or, worse still, to the ellipsoid. Thus the third dimension enters in triumph into the field of practical geodesy three centuries after it had entered with equal triumph into the realm of speculative geodesy through the works of Galileo, Kepler, and Newton.

There is no need for me to tell you how this fact leads to a radical revision of our way of looking at geodetic problems and of arriving at their solution.

What paths we shall have to follow we still do not know with certainty, although some of them have already been mapped out. This argument will in fact be dealt with next July in Venice, during the symposium organized for this precise purpose by our association.

For the present purpose it will be enough for me to draw attention to this turning point that we have now reached. So I come to the point of what I have said so far, namely that all this ties in with this occasion of the opening of our symposium. For we are in fact taking part in this triumphal entry of the third dimension in this first meeting to discuss the means of measuring distances in space directly, that is along geodesics of space, and not, as in classical geodesy, along geodesics on a surface.

Dazzled as we are by the marvelous technical apparatus which the latest methods for the direct measurement of distances place before our eyes, there is nevertheless the risk, owing to the way we are bound to this flat earth and bound, far more than we realize, to traditional ways of thinking, that the revolutionary aspect of the new instruments and methods may escape us. I am the first to admit that much time will pass before we can really get used to thinking and working in this new, vastly

extended world which is, even so, the natural domain of our science. In the same way I am convinced that we shall still, for a long time to come, have recourse to the fearsome bulldozer of the 'reductions' in order to attribute to the geoid, or to the ellipsoid, properties which really belong to the surrounding space, and which the instruments and methods that we have today before us will enable us to locate and describe.

But this time lag ought not to daunt us. Every conquest needs time to make itself felt, and I am quite certain that today's symposium marks a milestone in the study of geodesy, not only on account of the progress that it sanctions in the field of instrument development, but also and above all because it marks a new stage in the experimental conquest of physical space around us.

Atmospheric Limitations on Electronic Distance-Measuring Equipment

MOODY C. THOMPSON, JR., HARRIS B. JANES, AND FRANK E. FREETHEY

*National Bureau of Standards
Boulder, Colorado*

Abstract. In recent years various instruments and techniques have been developed for measuring distances electronically. The accuracy of such measurements depends on the accuracy with which time and the velocity of propagation of radio waves can be determined.

The National Bureau of Standards, under the sponsorship of the Air Force Ballistic Missile Division, has been studying the effects of atmospheric turbulence on the performance of radio distance- and/or velocity-measuring systems. Many of the data collected in this study are applicable to the problem of determining the accuracy of radio surveying methods. Long-term variations in the apparent length of a 15.5-mile path in Hawaii are shown, along with the effects of correcting for the atmospheric refractive index observed at 2 to 5 points along the path.

The development of electronic distance-measuring techniques during the past 20 years has resulted in radical changes in both the technical and the economic aspects of surveying. The measurement of base lines that formerly required many man-days or weeks can often be made now to the same precision by two men literally in a matter of minutes.

The geodesists are to be commended for their willingness to adopt these new methods and for the effective manner in which they have employed the equipment made available to them.

The physical basis for these techniques is essentially the simple expression that distance is the product of speed and time. More specifically, if we know the speed at which a radio or light signal travels through the atmosphere, and the time required for it to progress from one point to another, we can easily calculate the distance as the product of these two quantities. It follows immediately that the precision of this calculation is dependent upon the precision with which we can make these two independent determinations.

Each of these measurements involves an area of physics that is of importance to the National Bureau of Standards. The time measurement involves the sciences of electronics and time standards; the speed measurement is a problem fundamental to electromagnetic wave

propagation. Consequently, our work at the Bureau has led us to detailed consideration of these problems and particularly to a study of the factors that limit the precision of such measuring techniques. Although means may later be developed to circumvent what appear to be 'basic' limitations today, it is believed that some of the presently recognized factors will influence significantly the practical limits of precision for all the systems that have been publicly described to date.

Historically, the Radio Propagation Engineering Division of the Boulder Laboratories, National Bureau of Standards, has been conducting an experimental and theoretical study of the general problem since November 1954 [Herbstreit and Thompson, 1955]. During this time, the work has been supported almost completely by the United States Air Force and, because of the needs of the sponsor, has been directed almost exclusively at one aspect of the general problem, namely the specific question of the effect of atmospheric turbulence.

Because the lower atmosphere of the earth is generally turbulent, the density and composition of air vary, at every point, with time. Similarly, at any given time these physical characteristics vary from point to point. Since the speed of propagation of radio or optical signals through the atmosphere depends on the composition of the atmosphere, several effects

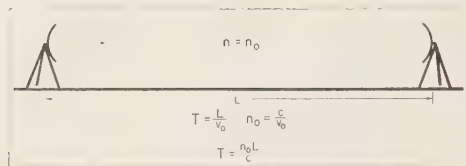


Fig. 1. Propagation through idealized homogeneous atmosphere.

are observed in actual atmospheric propagation paths that restrict our use of such signals. Accordingly, the Bureau's work in this area has been generally aimed at evaluating and, wherever possible, systematically classifying these atmospheric effects.

The following will be confined chiefly to a discussion of these atmospheric characteristics; that is, we will neglect the uncertainties in t in the equation $L = vt$, and assume that errors in L are due solely to errors in v . Now consider the physical situation shown in Figure 1. This is the idealized but trivial case in which the speed is a constant value v_0 at all points along the path. If we knew this situation to exist, we would need only to measure v_0 at any point in the medium to make the necessary conversion of our time measurement to distance. Furthermore, the error in our calculated distance would be determined by the precision with which we could measure v_0 .

In reality, this situation is not even approximated very often; we are, in general, faced with the case illustrated in Figure 2 in which, owing to the heterogeneity of atmospheric composition, the speed of propagation is a function of location. If we express this fact mathematically as $n = n(x)$, the actual conversion factor from our time measurement to distance should be the space average defined as

$$\bar{n} = \frac{1}{L} \int_0^L n(x) dx$$

Now, since we are always limited, physically, to determining $n(x)$ at some finite number of values of x , the above expression may be interpreted to mean that, even though we made errorless determinations of the index at each of k sampling points, our average would always be subject to a finite sampling error to the extent that

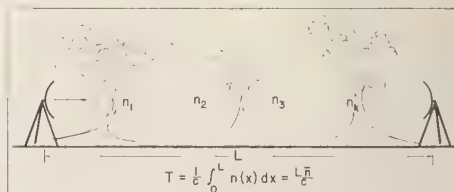


Fig. 2. Propagation through realistic turbulent atmosphere.

$$\frac{1}{k} (n_1 + n_2 + \dots + n_k) \neq \frac{1}{L} \int_0^L n(x) dx$$

This is the error that our experiments attempt to evaluate.

The early experiments were conducted in the Pikes Peak region of Colorado [Herbst and Thompson, 1955]. Subsequently, a series of measurements were made on the island of Maui, Hawaii [Thompson and others, 1959], and were at present set up in the Boulder area [Thompson and Janes, 1959].

The experimental approach has consisted of a rather straightforward process in which two radio instruments [Thompson and Vetter, 1958] operating on approximately 9400 Mc, were set up on rigid pedestals. The pedestals varied with the particular antennas being used but generally they were columns about 2 feet across and 2 to 5 feet high. Figure 3 shows 4-foot- and 18-inch-diameter dishes. The signals from the electronic time-measuring circuit were recorded continuously by means of both paper chart and magnetic tape recorders. In addition to the analysis of slow variations (such as diurnal), the power density spectrum was computed for variations in transit time up to the 10 cycles per second components.

During this period of radio observations psychrometer and barometer observations were made, from which the corresponding values of radio refractive index could be computed.

In the Maui experiments, recordings were made of temperature, pressure, and relative humidity at three intermediate stations in addition to the two ends of the radio path. The path geometry and the locations of these stations are shown in Figures 4 and 5. These data have been analyzed as follows.

At each hourly interval the variation in path



Fig. 3. Green Mountain mesa terminal.

length was calculated in two ways: first, by using the index measurements from only the two end points as the velocity correction, and then by using the data from all five meteorological stations weighted according to their differences in elevation. Figure 6 shows these two curves as well as the uncorrected variations. For each curve the standard deviation has also been computed.

The effects of averaging the observations to different degrees are illustrated in Figure 7. The ordinates in these curves are the running averages of the points plotted in Figure 6. The results of this process are summarized in Figure

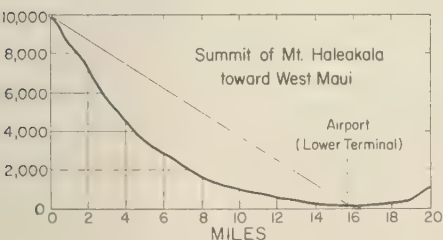


Fig. 4. Terrain profile for Maui phase measurements.

8, which shows the decrease in standard deviation as a function of the averaging interval for the different correction processes.

Although this work has been in progress for more than four years, only a small part of the data has been analyzed in the manner just illustrated. The primary interest of the United States Air Force as sponsor of this work has been, and is currently, with short-term vari-

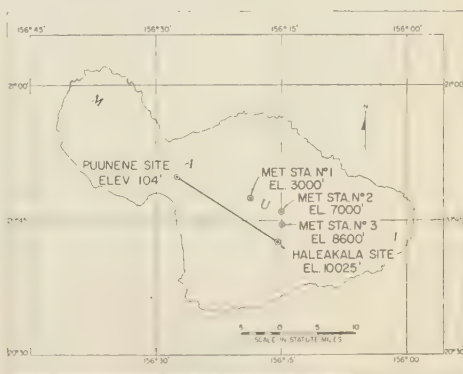


Fig. 5. Location of propagation path used in Maui experiment.

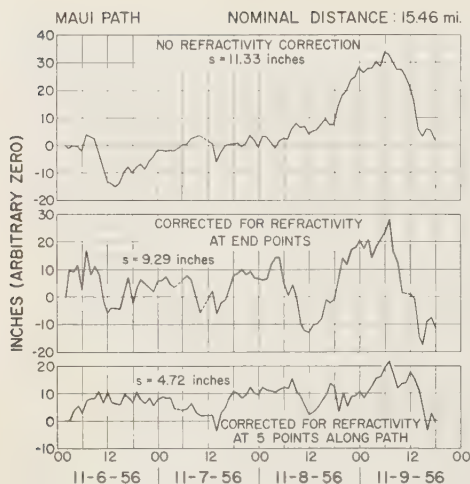


Fig. 6. Time variations in apparent distance.

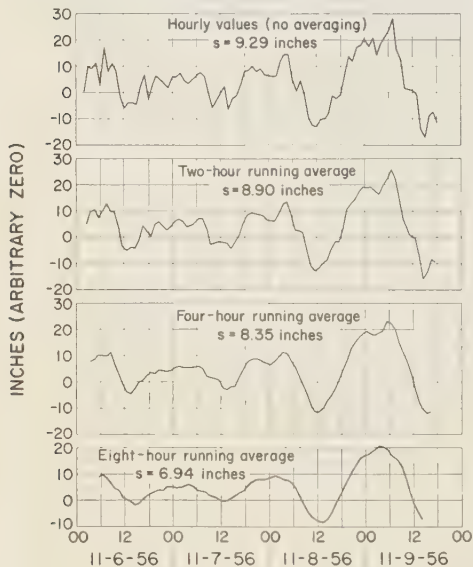


Fig. 7. Effect of averaging variations in apparent distance. Data corrected for refractivity at end points.

ability rather than with actual distance measurement. Consequently, the experiments have always been designed to emphasize the former while making no attempt to pursue the latter. Fortunately, a great many of the data contain

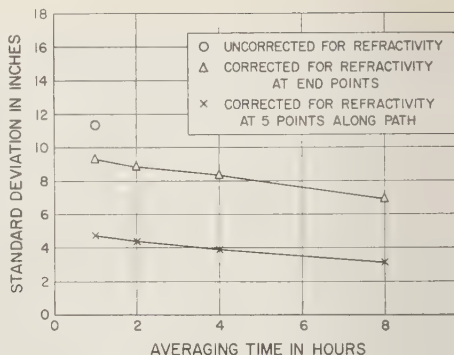


Fig. 8. Apparent distance variations versus averaging time. Maui path, Nov. 6-9, 1956.

information relevant to the latter application and may be of value in assessing the performance to be expected from distance-measuring systems.

The National Bureau of Standards' experiments have been designed from the beginning not simply to observe the turbulence effects but also to study them quantitatively. Thus, we were forced to develop instrumentation whose noise level was significantly lower than the variations in transit time resulting from turbulence.

One might say that our signal—to be studied quantitatively—is the electronic surveyor's noise level. The result has been equipment whose stability is such that variations of transit time of less than 1 micromicrosecond can be detected and the normal atmospheric variations can be recorded with an accuracy of a few per cent. At the same time, although for different purposes, we have had to develop instruments such as the National Bureau of Standards microwave refractometer, which can record variations in radio refractive index to a few parts in ten million.

In a homogeneous atmosphere, we are, though already instrumented to make electronic distance measurements accurate to the order of a few parts in ten million. Our experimental observations as described above, however, indicate that we are limited to accuracies of the order of several parts per million in the field.

It thus appears that we are clearly at the

nt where our lack of knowledge of atmospheric turbulence is the limiting factor and t, if we are to extend the precision of these etronic techniques, our first approach should to pursue this area further both experiment- y and theoretically.

REFERENCES

rbstreit, J. W., and M. C. Thompson, Jr., Measurements of the phase of radio waves received over transmission paths with electrical lengths varying as a result of atmospheric turbulence, *Proc. IRE*, 43 (10), October 1955.

Thompson, M. C., Jr., and H. B. Janes, Measurements of phase stability over a low-level tropospheric path, *J. Natl. Bur. Standards*, D, Radio Propagation, 63D (1), July-August 1959.

Thompson, M. C., Jr., H. B. Janes, and A. W. Kirkpatrick, An analysis of observed atmosphere-induced phase variations in line-of-sight microwave propagation, A paper presented by H. B. Janes at the Spring IRE-URSI Meeting in Washington, D. C., May 1959, unpublished.

Thompson, M. C., Jr., and M. J. Vetter, Single path phase measuring system for three-centimeter radio waves, *Rev. Sci. Instr.*, 29, (2), 148-150, 1958.

Physical Principles of the Electro-Optical Determination of Distances

A. KAROLUS

*University of Freiburg
Freiburg, Germany*

In principle an electro-optical distance-measuring device is identical to an apparatus to determine the velocity of light. The velocity of light is determined from the relation of the distance traversed by light and the time it requires to do so. The transit time of the light in the methods described below is equal to the period of a usually sinusoidal modulation of the light, as a result of which the light emanating from the transmitter of the distance-measuring device has either the same phase as that of the light returning over the distance of measurement or the opposite phase.

The modulated light traverses a semitransparent mirror, and a portion of it enters a photocell within the optical system of the transmitter and there gives rise to an alternating current having the same phase as the light emanated. The remaining light crosses the measurement path, at the end of which it is reflected and gives rise to an alternating current in a second photocell in the receiver. The phase of this current generally differs from that of the current in the photocell of the transmitter. For a known velocity of light and a known frequency of modulation a phase meter can be immediately calibrated in distances. If the phase meter can determine only certain phase displacements, for example the angle 0° or 180° , the frequency ought to be varied until this phase angle is attained.

The voltage for the light source or the light modulator is obtained from a high-frequency generator whose frequency is determined by a frequency meter or a similar device.

The requirements placed on the precision of an electro-optical distance meter are varied, according to the application intended. In surveying a base the geodesist requires a precision of $1:10^5$; very often an accuracy of $1:10^4$ or

less suffices. In calibrating an electro-optical distance meter the velocity of light must be known. The velocity of light is still uncertain to 1–3 km/sec, i.e., approximately $1:10^5$; therefore the absolute error of an electro-optical distance meter cannot be less than $1:10^5$.

Two methods for the determination of c will be discussed, the principle of which has been employed in the distance meters developed during the present time.

THE METHODS OF FIZEAU AND OF KAROLUS

In the classical method of Fizeau the image of a point source of light is formed on a toothed wheel, and its intensity is modulated by the rotation of the wheel. The modulated light traverses a known distance, at the end of which it is reflected by a mirror and passes the rotating toothed wheel a second time. A portion of the light is observed through a semitransparent mirror. The brightness of the source seen by the observer depends on the phase of the reflected light in reference to the position of the cogwheel. If the reflected light impinges on the toothed wheel at a time such that it passes through a gap through which the transmitted light passes, the cogwheel has been replaced by a tooth, the light beam or the observed visual brightness is at a minimum. For a constant distance and a constant frequency of modulation, during the traversal the observer registers alternating maxima and minima in the brightness, at an increasing number of revolutions per unit time. Thus in Fizeau's method the rotating toothed wheel is not only the light modulator but also the phase meter (synchronized shutter).

If the first minimum is used for measurement, the phase of the returning light can be determined to only about 1 per cent of a period. This inaccuracy in the phase determination remains unchanged even if the total phase d

ment of the light is a multiple of 180° at a higher frequency of modulation. On the other hand, the relative error in the distance determination decreases in proportion to n with increasing order number. Since the mechanical cogwheel allows only about 10^4 light interruptions per second, the distance of measurement was enlarged to approximately 50 km ($n = 8$) by Fizeau's successors.

Proceeding from this arrangement of Fizeau, the author and his co-workers carried out a series of c determinations, during 1925-1940, in which the rotating cogwheel was replaced by a noninertial method of light modulation. By means of the electro-optical Kerr effect a light beam can be modulated up to a frequency of 10 Mc/s. A modulation frequency of 10 Mc/s corresponds to an electrical wavelength or light path of 30 meters. The first minimum occurs at a light path of 15 meters, i.e. at a separation of 7.5 meters between mirror and transmitter.

The phase of the reflected light could be observed through the same Kerr cell used for the modulation, but for optical reasons it is better to observe the reflected light through a second Kerr cell. This second cell is electrically in parallel with the modulating cell so that the light transmissibility varies in phase with both.

Since the light passes through the second cell with a time displacement, maxima and minima in the light intensity occur periodically by varying the frequency or the light path despite the electrical phase equivalence of the two cells.

LIGHT SOURCES FOR THE ELECTRO-OPTICAL DISTANCE METER

The image of the light source formed by the transmitting optics over the reflecting mirror at the lens or the cavity mirror of the receiver is always larger than the receiving lens itself, under the conditions of electro-optical distance determination. Under these circumstances only the surface intensity of the light source and the diameter of the transmitting and receiving optical systems are contained in the transmitted light beam. The light current ϕ is given by

$$\phi = 1.53 \cdot 10^{-3} \cdot A \cdot B \cdot D^4$$

$$\cdot (1/E^2) \cdot e^{-8E/S} \quad \text{lumens}$$

where A is a constant of the apparatus, which takes into account losses in the optics and in the light modulator; B is the surface brightness of the light source in Hefner candles per square centimeter; D is the diameter of the objective in meters; E is the distance to be measured in kilometers; and S is the normal seeing range during the course of the measurement.

In Table 1 several light sources suitable for electro-optical distance determinations have been listed with information about their surface brightness and their modulating ability. The light source employed should have as high a surface brightness as possible; it is even more important, however, that the surface brightness not change with position and that the noise of the light be determined only by the number of quanta and not the nature of the light excitation.

Phosphors excited by electrons of a sufficiently high velocity have a surface brightness of 10^4 to 10^5 Hefner candles per square centimeter in the event that the power transformed at the focal point attains the order of magnitude of 1 watt. Such a power would destroy the phosphor within a fraction of a second if the electron beam remained focused on it. The light source of an electro-optical distance meter must remain at the same spatial position. Suitable solutions to this difficulty are presented by Karolus [1958], as are also the advantages and disadvantages of the various light sources contained in Table 1.

LIGHT MODULATION

The light modulation of an electro-optical distance meter must fulfill the following conditions: (1) all bundles of light must be modulated in the same phase; (2) the degree of modulation of all bundles must be the same; (3) the degree of modulation must be as large as possible, and the distortion as small as possible; (4) the degree of modulation must be constant in the frequency region employed; (5) the energy required to regulate the light should only be a few watts.

TABLE 1

Source of Light	Brilliance, Hefner candles/cm ²	Electric Power, watts	Frequency to Which Source Can Be Modulated cps
Incandescent lamp	800-1200	2-30	100
Carbon arc	14,000-18,000	500	1000
Zircon arc	4000-8000	20-100	1000
High-pressure mercury arc	10 ⁴ -10 ⁶	100-500	10 ⁶
Phosphors	10 ⁴ -10 ⁶	1-2	10 ⁷
Spark	ca. 10 ⁶	2-5	10 ⁶

If distances of 1 to 30 km are to be measured electro-optically with a precision of $1:10^6$, the light must be modulated in the frequency region 10^5 to 10^7 cps, in view of the error in the phase determination. Several light sources whose radiation can be directly modulated will now be discussed.

DIRECT MODULATION OF THE LIGHT SOURCE

Arcs. Arcs in mercury vapor and in noble gases are characterized by a high surface brightness (cf. Table 1). A high-pressure mercury lamp with a power rating of 200 watts has a potential drop of about 30 volts at a vapor pressure of 50 atmospheres and an arc length of 1.3 mm. If an alternating current is superposed on the direct current of the arc, the light of the lamp is modulated. The degree of modulation remains constant for frequencies between 50 cps and 25 kc and corresponds to the value of the current modulation. In the region from 25 to 100 kc, on the other hand, the light modulation decreases to about 35 per cent.

Unfortunately the distribution of the light modulation along the discharge is very non-homogeneous. Next to the electrodes the surface brightness and degree of modulation are essentially greater than in the center of the arc. Moreover, various regions of the arc display considerable differences in the angle of phase. For these reasons the directly modulated high-pressure mercury arc is unsuitable in an electro-optical distance meter. On the other hand,

it can be employed as a light source which can be driven by direct current and whose light beam is modulated by a separate control mechanism.

Phosphors. The light intensity of a phosphor excited by cathode rays can be varied by modulating the electron ray. Table 2 lists several suitable phosphors, their color of light and their afterglow.

Of all known phosphors emitting in the visible region of the spectrum, gehlenite has the smallest afterglow constant. On a tube having an anode voltage of 25 kv and a beam current of 10 microampere we measured the following values: the degree of modulation of the light amounts to 90 per cent at 1 Mc/s and decreases to about 50 per cent at 10 Mc/s.

Should it be possible to develop a cathode ray tube with gehlenite whose energy absorption amounts to 0.3-1 watt and whose light point remains stationary in space, such a tube would be superior to every other light source for electro-optical distance determinations.

INDIRECT LIGHT MODULATION

Table 3 gives several methods of indirect light regulation.

The Kerr effect. Certain liquids become birefringent on the application of an electric field. If linearly polarized light, the vibration plane of which is oriented at an angle of 45° to the field direction, passes through such a liquid the components of the light vector parallel and perpendicular to the field undergo a phase displacement. The linearly polarized light entering the liquid in general comes out elliptically polarized. If the path difference of both components of the light is a half wavelength, the

TABLE 2

Phosphor	Chemical Composition	Spectral Maximum, Å	Afterglow sec
Calcium tungstate	$\text{CaWO}_4 : \text{W}$	4300	6-12 · 10 ⁻⁶
Zinc sulfide	$\text{ZnS} : \text{Ag}$	4400	3-10 · 10 ⁻⁶
Zinc oxide	$\text{ZnO} : \text{Zn}$	5050	3-10 · 10 ⁻⁶
Gehlenite	$2 \text{CaO} \cdot \text{Al}_2\text{O}_3 \cdot \text{SiO}_2 : 0.015 \text{Ce}_2\text{O}_3$	4000	3 · 10 ⁻⁶

TABLE 3

Effect Employed	Frequency Range, cps	Control Voltage Required
Kerr effect: electrical birefringence in certain liquids	0-10 ⁷	3-8 kv _m and 1-2 kv _{eff}
Electro-optical birefringence in crystal such as ammonium phosphate	0-10 ⁸	2 kv _{eff}
Birefringence at vibrating quartz crystals	10 ⁵ -10 ⁶ at definite frequencies only	Several 100 v _{eff}
Refraction by ultrasonics:		
(a) Stationary waves	10 ⁶ -10 ⁷	About 100 v _{eff}
(b) Modulated progressive waves	0-2.10 ⁴	About 100 v _{eff}
Refraction by stationary sound waves in solids	About 10 ⁷ at definite frequencies only	Several 100 v _{eff}

ht coming out is again linearly polarized and a completely pass through the analyzer.

Of all the materials known at present, nitrobenzol is best suited for a technical light shutter. It has a very large Kerr constant, high breakdown voltage, low light absorption, high specific resistance, and low dielectric losses.

On the application of an electrical field the birefringency arises in less than 10⁻¹⁰ second. Once the field has been removed it disappears as a result of thermal motion in a similarly short time.

After repeated vacuum distillation the specific resistance of nitrobenzol amounts to about 10¹⁰ ohm cm. Immediately on the application of a d-c voltage the current is 100 times the initial value, even in carefully purified nitrobenzol. During the first few minutes the electric field causes an ordering in the distribution of conducting and nonconducting parts of the nitrobenzol. A simultaneous stratification of the field is connected with this ordering. The field strength directly at the electrodes is essentially greater than in the center of the field. Every one of the Kerr cell has a different birefringency and hence a different degree of modulation. The layering becomes more pronounced the smaller the electrode separation of a Kerr cell. A variation in the phase of the light modulation occurs parallel to the ordering. In the extreme case the light in the boundary zone can be modulated in phase opposition to the light in the center of the field. These positional phase deviations make it necessary to blend out the boundary zones and to use only the ap-

proximately homogeneous central part of the Kerr cell to regulate the light.

The electro-optical birefringency of crystals. Crystals like quartz, tourmaline, potassium, and ammonium phosphate also become birefringent in an electrical field. Contrary to the Kerr effect they also exhibit an electro-optical effect in this case, if the light extends in the field direction. The frequency characteristic of the birefringency varies unexpectedly. The light modulation in the region from 50 cps to 20 kc/s is practically constant for a crystal of the dimensions 30 by 30 by 5 mm. It has a maximum at both 39 and 90 kc/s, the height of which depends only on the damping of the respective crystal. By shortening the sides of the rectangular crystal the resonance points can be displaced to higher frequencies. The characteristic vibrations of a crystal of sides 21 by 21 mm lie at 56 and 130 kc/s. These mechanical resonances can be suppressed by sufficient damping.

The modulation of all ammonium or potassium phosphate crystals investigated decreased to 25 per cent at 200 kc/s and 10 per cent at 3 Mc/s. From this it must be concluded that the noninertial component of the electrical birefringence, the true Kerr effect, is less than 10 per cent, the greater part of the electro-optical effect for phosphate crystals being due to the mechanical deformation of the crystal in an electrical field.

To modulate traversing light the birefringency of quartz crystals excited at their characteristic vibrations has sufficiently high values.

Ultrasonics. A plane sound wave generated in a liquid refracts light just like an optical grating. This refraction was suggested as a means of modulating light by the author in 1932.

A quartz plate cut in the x plane is placed in a glass vessel filled with carbon tetrachloride or paraffin oil. The transverse vibration of the plate is excited by an alternating electrical field. At a sound velocity of 1000 m/sec in the liquid and at a characteristic vibration of the quartz of 10^7 cps, the wavelength of the sound wave is 0.1 mm.

The image of a line source of light, for example a straight tungsten wire 0.2 mm in diameter, is cast on a slit. All the light passes through the slit for nonvibrating quartz. If an ultrasonic grating is made, refractive images occur symmetrically to the original position of the image of the light source. If the slit is arranged so that the zero order (the unrefracted light) is passed through, the zero-order light decreases as the amplitude of the sound wave increases and is wholly refracted in the higher orders at a definite alternating potential on the quartz.

If a wire or a band of such width that the zeroth order is just blended out is substituted for the slit, all the refracted light reaches the following optical system. The light current increases as the tension on the quartz (i.e., the voltage) increases.

The light is refracted in the same manner for progressive as for stationary sound waves. For stationary sound waves the refracted light is modulated at twice the frequency of the alternating voltage on the quartz. For a progressive wave of constant amplitude, however, the refracted light remains unmodulated.

STATIONARY SOUND WAVES

If a plane reflector is placed opposite the quartz at a distance that is an exact multiple of a half wavelength, standing sound waves arise. The nodes and loops remain fixed in the liquid. The refracting grating arises and disappears at every half period of the quartz vibration, and the total refracted light is consequently high-frequency-modulated. It has the same phase at all points in the sound field.

The creation of standing sound waves in-

volves many technical difficulties. The heating of the liquid by the sound vibrations changes the temperature-dependent sound velocity. In addition, as a result of the damping caused by every liquid, the progressive and reflected waves are of different amplitudes at the same point and thus a progressive wave superposes itself on the stationary sound wave, causing phase differences in the light modulation of different zones of the sound field.

An arrangement due to Bär circumvents the difficulties arising from the adjustment of the reflector. In the parallel radiation path between the two lenses two quartz crystals of the same frequency are so arranged in the direction of the light that their sound fields are oppositely directed. Absorbers are placed opposite both quartz crystals to suppress every reflection. The two sound waves traveling in opposite directions are identical to a stationary sound wave in their refractivity.

Standing sound waves can also be excited in uniformly bounded solid bodies, as for example in quartz crystals or in glass cubes, when their characteristic vibrations are excited. Here also the unrefracted light is modulated at a frequency twice the frequency of the exciting vibration. A reflector is not necessary, yet only definite frequencies can be excited, namely the elastic characteristic vibrations of the glass or quartz cube.

PROGRESSIVE WAVES IN LIQUIDS

The amplitude of light refraction with a progressive sound wave can be modulated by superposing a signal voltage on the high-frequency alternating voltage which excites the characteristic vibration of the quartz crystal. The characteristic frequency of the quartz crystal must be at least an order of magnitude greater than the highest modulation frequency.

Because the velocity of sound is finite (900 to 1800 m/sec) the spatial phase differences of the low-frequency pressure distribution are large. Even if the refraction is confined to a zone of 2-mm width the phase differences of the individual light bundles still amount to 40° – 80° . If a part of the modulated light is cut off by a distant mirror or in the receiving optics, the phase of the remaining light and thus the distance measured change. This error

must be eliminated by optical or electrical means.

FREQUENCY MEASUREMENT

In all methods of electro-optical distance determination the light is modulated sinusoidally at frequencies in the range of 100 ks/s to 10 Mc/s. If the distance-determining device is to have an accuracy of $1 : 10^5$ the modulation frequency must be determined to an accuracy of $1 : 10^6$. Every frequency determination is reduced to a comparison of the unknown frequency with that of a quartz standard or with a frequency derived from it. The quartz reference has a frequency of 10^5 or 10^6 cps. At constant temperature the frequency of a quartz crystal varies by less than $1 : 10^7$ over a period of a few days.

If the electro-optical distance meter operates at a set modulation frequency, the light modulator is regulated directly with a vibrating quartz crystal or a harmonic of it. If the modulation frequency of the light is variable it can be accurately determined to 1 cps with an electronic counter at a measuring time of 1 second.

LIGHT MEASUREMENT: PHOTOCELL AND PHOTOMULTIPLIER

The electro-optical measurement of a distance consists in the determination of the phase difference between the modulated light emanating from the transmitting optical system and the light re-entering the receiving optical system. If the phase meter is electrical, the light is first transformed in a photocell, or in a photomultiplier, into an alternating current of the same phase. The difficulties arising thereby lie in the transit time of the electrons and in the spread in transit time. The transit time is a function of the voltage on the photocell and of the distribution of the light on the photocathode. The distance corresponding to the transit time must be added to the distance to be determined. The uncertainty in determining the distance equivalent to the transit time arises from the path differences of the individual photoelectrons or of the secondary electrons emitted.

For an emf of about 100 volts the electron velocity is 1 per cent of the velocity of light.

The equivalent light paths are accordingly 100 times as long as the electron paths.

A photocurrent can be amplified by the repeated generation of secondary electrons. According to the voltage applied, a ten-stage multiplier can attain an amplification of 10^6 to 10^8 . The sensitivity of such a multiplier, that is, its output current per lumen, increases with the amplification and amounts to 10^2 to 10^4 amp/lumen. The amplification depends strongly on the over-all voltage. Doubling the voltage increases the amplification by 10^4 .

The transit time and its spread increase with the number of multiplying stages (the number of dynodes). For various ten-stage photomultipliers with an over-all potential difference of 1500 volts the transit time lies between 4 and $8 \cdot 10^{-8}$ second, which corresponds to a light path of 12 to 24 meters. If the transit time were invariant, the equivalent light path of a multiplier for a definite anode voltage could be considered. But if the voltage varies by 1 per cent an uncertainty in the equivalent light path of 6 to 12 cm arises.

In an electro-optical distance meter the light incident on the cathode constantly changes its position because of air disturbances and the movement of the reflecting mirror. This movement of the receiving light on the photocell causes additional transit time differences of $1\text{--}4 \cdot 10^{-9}$ second, which correspond to path differences of 30 to 120 cm.

An uncertainty of 100 cm in a distance of 10 km causes an inaccuracy of $1 : 10^4$. If the accuracy of measurement is to be greater than this the photomultiplier is unsuitable.

In recent years the author has developed a photocell of small transit time and extremely small transit time spread. The cathode and anode are placed on plane glass or metal surfaces which are as parallel as possible at a separation of 1 to 2 mm. For a direct voltage of 100 volts the electron transit time in the photocell is $5 \cdot 10^{-10}$ second. If the anode were perfectly parallel to the cathode, the transit time spread would be zero. In the cells constructed till now a deviation in parallelness of 0.1 mm on the average has been attained. A displacement of 5 mm in the emission area results in a variation in the transit time of $1 \cdot 10^{-11}$ second (equivalent light path = 3 mm).

Photocells having symmetrically arranged electrodes display an effect at alternating voltages greater than 1 Mc/s which is known as dynamic multiplication. It attains its maximum value at a frequency for which the time of a half period is equal to the electron transit time.

In this region the photocells are unsuitable for light and phase measurements. In the newly developed photocell the multiplication first occurs at about 200 Mc/s because of the small cathode-anode separation.

PHASE MEASUREMENT

Synchronized shutter with visual measurement of the brightness of the returning light. A phase measurement must be independent of the intensity of the measuring light. Even variations of 1 : 10 in the light may not have an effect on the indicated phase greater than the admissible uncertainty. In the classical method of Fizeau the rotating cogwheel served to modulate the outgoing light and to determine the phase of the returning light. It has already been said that the accuracy of the visual determination of the first light minimum was at best 1 per cent.

A synchronous shutter is suitable as a phase meter only if it has the same transmissibility at all points within a certain instant, i.e., if no positional phase differences arise, as in the light modulator. In the cog disk of Fizeau the border zones of the light-source image formed in the plane of the wheel are released at different times. Thus the cogwheel is equally unsuitable as a modulator or as a phase meter.

In comparison with the cogwheel the phase errors of a Kerr cell are 100 times smaller.

In the method of crossed cells (Karolus and Mittelstädt, 1928) variations in the light source and changes in the light along the path cause no measurable errors.

Synchronous shutter with objective determination of the light minimum. The light minimum behind a synchronous shutter can be more accurately determined by an objective intensity measurement than by visual observation. A selenium cell or a photocell with a secondary electron multiplier measures only the amount of the light current, not its phase. The transit time errors of a photomultiplier, then,

have no effect on the accuracy of the measurement.

Vector addition of two modulated light beams. In the arrangement for the determination of c due to Anderson, amplitude-modulated light from two different mirror positions thus having a definite phase difference, is simultaneously incident on a photocell. A voltage is applied to the photo cell, and a resonant circuit tuned to the modulation frequency is placed in the anode current. The a-voltage in the resonant circuit is a maximum if both the simultaneously incident light beams have the same phase; it is a minimum for phase opposition. Actually, the photoelectron emitted because of the light beams, and not the two light beams, add vectorially. The resulting light beam depends on the distribution of the emission of the photocathode and on the transit time differences of the electrons emitted at different points on the cathode.

Anderson's method yields accurate results as long as the intensities of the two light beams and their distribution on the cathode do not change. The method is unsuitable as an electronic optical distance meter, for the intensity of the measurement light varies constantly and thus comparison of the amplitudes of the light quantities is not possible.

The photocell as a phase-dependent rectifier. The phase measurement suggested by the author in 1935 and carried out in 1936 to 1939 by A. Hüttel differs from the Anderson method basically in that an a-c voltage is applied to the photocell, which has a frequency equal to that of the light modulation. A portion of the modulation a-c voltage of the Kerr cell is on the photocell. Because the Kerr effect is noninertial the a-c voltage on the photocell is in phase with the emanating light of the arrangement.

For a constant amount of modulated light the arithmetic mean of the photocurrent is a maximum if the returning light is in phase with the a-c voltage on the photocell. From the distance between two points having the same phase—the same mean photocurrent—the wavelength is obtained, and, for known frequency, the distance is obtained from the wavelength.

Before every phase measurement, the light must be adjusted to the same intensity inde-

endently of the phase. To check this, a d-c voltage of 200 volts is applied to the photocell and the magnitude of the light beam is kept constant by a diaphragm or a gray wedge.

In 1938 the author suggested that the alternating voltage on the photocell should be periodically reversed by 180° . The mean photocurrent remains unchanged on reversing the phase if a difference of 90° exists between the light and the a-c voltage. For all light paths corresponding to a quarter wavelength or to a multiple of $\lambda/4$, all a-c components of the switching frequency (normally 25 cps) disappear in a low-frequency amplifier inserted after the photoresistance.

In experiments carried out by H. Mende at the suggestion of the author in 1938 to 1940, the accuracy of measurement of the first order amounted to $1:10^4$. Variations in the light source and in the absorption of the light along the path result in no errors in measurement.

In 1941 to 1949 E. Bergstrand developed an electro-optical distance meter in which the light is modulated by means of a Kerr cell as in the determinations by Karolus and Mittelstädt, and the phase is measured with a photocell or a photomultiplier with an a-c voltage applied, analogous to the Karolus-Hüttel method. The new feature in Bergstrand's arrangement is that the light modulated in the Kerr cell changes its phase periodically by 180° . Instead of the usual d-c bias, a line frequency rectangular a-c voltage of maximum value 5 kv is on the Kerr cell.

The light path for which the switching frequency of 50 cps in the light beam disappears is determined. For a light modulation of 10 Mc/s the null condition is first fulfilled for a light path of 7.5 meters, hence a mirror separation of 3.75 meters. In general the zero adjustment is made by moving the more distant reflecting mirror.

If the mirror cannot be positioned exactly, Bergstrand varies the frequency of light modulation by relatively small quantities, or he enhances the light path within the optical arrangement until the zero condition is fulfilled.

As receiving photocell Bergstrand employs a nine-stage photomultiplier and introduces the a-c voltage between anode and the last dynode.

All transit time errors of the multiplier enter fully into the measurement.

The electrical measurement of certain phases ($\varphi = 90^\circ, 180^\circ$, or 360°). The last-discussed methods of phase determination (Karolus-Mende, Bergstrand) by a periodic reversal of the a-c voltage on the phase-measuring photocell or the phase of the light beam are examples of the measurement of a definite phase. Only if there is a phase displacement of $\pm 90^\circ$ between the sinusoidal light beam and the a-c voltage does the mean value of the light current remain unchanged during the polarity reversal.

Phase displacements of 180° or 360° are measured either by the coincidences or anti-coincidences of pulses derived from the sinusoidal light beams. Pulses or sinusoidal a-c voltages may be measurably displaced by electrical delay lines or acoustical conductors. In general, however, such arrangements do not attain the accuracy required of electro-optical distance meters.

The electrical measurement of all phase angles. Among the numerous methods of phase measurement one developed by Kretzmer is discussed at length in Karolus [1958]. Both sinusoidal input voltages are transformed into square wave pulses by a multistage limiter. Pulses derived from these rectangles by differentiation control a bistable multivibrator. The mean value of the anode current of either of the two multivibrator tubes is proportional to the phase difference between the two signals. A coil instrument can be directly calibrated in phase angles or in distances.

The accuracy of the phase meter described is limited by the accuracy with which the coil instrument can be read. If a compensational method is substituted for the d-c measurement, the phase angle can be obtained accurate to $1:10^4$, i.e. to a few minutes of angle.

Transposition of the phase measurement into another frequency range. If the light beam is modulated at a high frequency, the transposition of the phase measurement into the frequency range near 1000 cps can have several advantages. The errors in measurement of coincidence circuits and of direct phase meters are smaller in the low-frequency region, and

the amplification and limiting are essentially simpler.

The disadvantage of transportation are that the frequency of measurement and the heterodyning frequency must be quartz-stabilized and that the method of measurement must be adapted to this fixed frequency of modulation.

The phase displacement of the two output signals remains constant in the beat frequencies created if the heterodyning signal is the same for both. To distinguish the high frequency and the difference frequency a resonant circuit tuned to them is necessary. Every fluctuation in the difference frequency causes a phase change in the lower frequency resonant circuit.

USEFUL SIGNALS, NOISE SOURCES, AND THE SIGNAL-TO-NOISE RATIO

The phase difference between periodic events of the same frequency is defined by the separation of the zero points. These separations can be measured as accurately as desired as long as the vibrations are undisturbed, i.e. as long as the distortion is small and the noise can be neglected. For a useful signal of the magnitude S and a noise signal N , the ratio S/N gives the limit of accuracy of the phase measurement.

The useful signal of an electro-optical distance meter is given by the magnitude and the degree of modulation of the photocurrent in the receiver. This, in turn, is proportional to the useful light incident on the photocell and to the sensitivity of the photolayer (micro-amperes per lumen).

The light current at the exit of the transmitting optical system amounts to about 0.01 lumen for a luminous surface of 0.3 by 0.3 mm and $B = 1400$ (an incandescent lamp). A light current of this magnitude causes a photocurrent of $0.5 \mu\text{a}$ in a photocell with a sensitivity of $50 \mu\text{a/lumen}$. A receiving mirror of 25-cm diameter and a focal distance of 2 meters receives only $1/36$ of the light current emitted across a distance of 5 km. As a result of absorption and scattering only about $1/100$ of the light actually returns to the receiver at ordinary conditions of sight. The current in the receiving photocell thus amounts to $1 \cdot 10^{-8}$ to $1 \cdot 10^{-10}$ ampere depending on the amount of absorption. In a photomultiplier this current

can easily be amplified to 1 milliampere but includes all the errors of the multiplier.

In a simple photocell a photocurrent of $1 \cdot 10^{-9}$ ampere gives rise to a useful signal $U_s = 1 \cdot 10^{-9} \cdot 10^7 = 10 \text{ mv}$ for a photoresistance of 10 megohms.

The thermal noise voltage on a resistance of $R\Omega$ at room temperature has the following magnitude:

$$U_R = 1.26 \cdot 10^{-10} \sqrt{R \Delta f} \text{ volt}$$

The band width Δf is equal to the reciprocal of the time constant of the measuring instrument or to the frequency of the switching for periodic changing the phase. If $\Delta f = 10 \text{ cps}$ the noise voltage becomes

$$U_R = 1.3 \cdot 10^{-9} \text{ volt}$$

and

$$S/N = 10^{-2} / (1.3 \cdot 10^{-9}) \sim 8000$$

The noise of the amplified photocurrent (shot effect) is given by the relation

$$J_{sh} = 5.6 \cdot 10^{-10} \sqrt{J_{ph} \Delta f} \text{ amp}$$

For $J_{ph} = 1 \cdot 10^{-8}$ ampere and $10^7 \Omega$ this shot portion causes a noise voltage of 0.6×10^{-9} volt, which gives $1.4 \cdot 10^{-8}$ volt when added quadratically to the thermal noise of the resistance.

The receiving optics invariably accept scattered light and also from an image of the background of the distant mirror on the photocell. By daylight this disturbing light can be a multiple of the useful light, as a result of which the shot noise increases correspondingly.

S/N can be increased by employing a light source having a high surface brightness, by decreasing the band width, by increasing the working resistance, and also by employing transmitting and receiving mirrors that are as large as possible.

CONCLUDING REMARKS

This discussion is restricted to the physical problems of electro-optical distance determina-

, and is concerned primarily with the quantitative evaluation of the systematic errors in methods of modulating light and of light phase measurement. Detailed studies are necessary if the electro-optical distance meter is to be sufficiently accurate for geodetic applications.

REFERENCES

- Karolus, A., Die physikalischen Grundlagen der elektrooptischen Entfernungsmessung, *Abhandl. bayer. Akad. Wiss.*, 92, 1958.
- Karolus, A., and O. Mittelstädt, Die Bestimmung der Lichtgeschwindigkeit unter Verwendung des elektrooptischen Kerr-Effektes, *Physik. Z.*, 29, 698-702, 1928.

The Geodimeter System A Short Discussion of Its Principal Function and Future Development

ERIK BERGSTRAND

*The Geodimeter Co.
Division of Berg, Hedstrom, & Co., Inc.
1170 Broadway, New York 1, N. Y.*

The basic principal of the geodetic distance meter called the Geodimeter is to measure the time for light to cover the required distance. Knowing the velocity of light, we get the distance. Before we can measure along the light beam it must be marked in some way. This is done by the aid of the Kerr cell, a glass vessel containing the highly purified fluid nitrobenzene and electrodes for applying a voltage to the fluid. The electrical field between the electrodes orientates the dipole molecules of the nitrobenzene in the direction of the field. The degree of orientation is dependent on the magnitude of the applied voltage. Simultaneously with the application of the voltage and the orientation of the molecules the optical properties of the nitrobenzene will be changed to those of a doubly refracting crystal. The optic voltage dependence of the Kerr cell is used to vary the intensity of the transmitted light. A direct voltage applied to the cell reduces the light intensity by half. The alternating voltage of a radio transmitter is also applied, which alternately increases and decreases the voltage and thus controls the light intensity from minimum to maximum. If the polarity of the direct voltage is changed, the radiofrequent sequence light, dark, light . . . etc., changes to dark, light, dark . . . etc. The feature of the Kerr cell is its ability to follow the rapid variation of the radio voltage.

The Kerr cell with light source and radio transmitter is the transmitting part of the Geodimeter. The receiving body is a phototube of multiplier type. It will emit electric current, photocurrent, when light falls upon its cathode and a working voltage of correct polarity is applied. If the poles are reversed there will be

no current, whether there is light on or not. The voltage, in fact, continually changes polarity, because it is the alternating voltage from the radio transmitter feeding the Kerr cell. The net result is that the sensitivity of the phototube alternates with the same frequency as that at which the Kerr cell acts.

Figure 1 shows the radio transmitter RT controlling the light intensity of the source L which emits rapid flashes. The flashes cover the distance to be measured to the reflecting system M and back again to the phototube P . In the instants of high sensitivity of the tube are simultaneous with the arriving flashes, we get

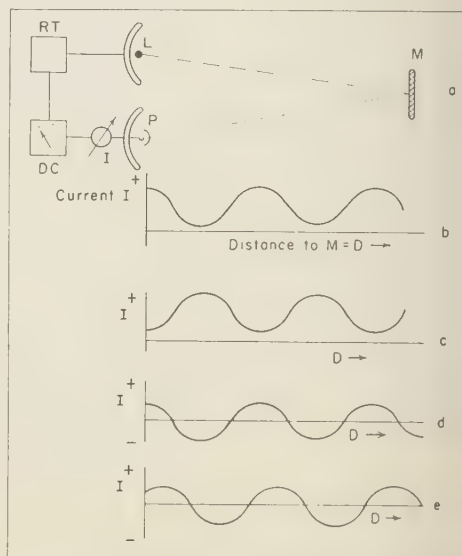


Fig. 1. A Geodimeter.

ing current through P as read off on the instrument I . When the reflector is a little farther away, the path for the flashes is longer and they arrive a little later, perhaps simultaneously with the insensitive moments of the meter; then there will only be a weak current to read off on I . Thus the strength of the current depends on the distance between Geodimeter and reflector. Figure 1*b* is drawn for the case when the sensitive phases of the phototube and the emitted light maxima coincide. Then if the light goes directly onto the phototube the current is maximum.

If the polarity of the direct voltage on the Kerr cell is changed we get the reverse state: the insensitive phases of the phototube coincide with the light maxima. In this case the current is like Figure 1*c*, and we start with a minimum of current close to the Geodimeter. Neither of these currents is suitable for a definition of the position of the reflector. For instance, at maximum and minimum points the strength of the current change is very small for a change of distance. The Geodimeter emits both these currents, however, because 100 times a second the polarity of the direct voltage on the Kerr cell changes sign. The two currents pass the instrument I in opposite directions, so that we can take the difference between them as in Figure 1*d*. Here, when the current goes through zero, we can use a sensitive instrument, and the change of current is large for a change in distance. The position of the reflector will be well defined. The distance between two consecutive zero positions depends on the rapidity of the flash emission, that is, on the frequency of the radio transmitter. Ten million flashes per second corresponds to the distance $7\frac{1}{2}$ meters in this case. A frequency of 10 Mc/s produces a radio wavelength of 30 meters. Light, with the speed of radio waves, goes to and fro, and thus the period of change corresponds to a distance of 30 meters, and two zero points in a period gives 15 meters. In the field, the reflector surely is not in a position of zero current; it may be anywhere between the positions of zero current. Now, if we let the feeding voltage from the transmitter RT to P pass through some coils of wire, the alternating voltage will be delayed; that is, for zero condition the light flashes must be delayed equally long, and we

assume the delay to be caused by this extra distance. The delaying coils can be chosen to move the zero point to coincide with the actual position of the reflector. Then the current should be as in Figure 1*e* as a result of the hypothetical moving of the reflector the whole path up to the actual position. The required distance, then, is an even multiple of half the unit length 15 meters, plus the distance to the first zero point, depending on the actual electric delay. This small excess is determined as follows: the electric delay remaining fixed, the light to the reflector is shifted over to a variable distance inside the Geodimeter called the light conductor. It is once for all calibrated for distance from some stipulated mark on the Geodimeter and is then adjusted in length to fit the delay just used for the distant reflector. The procedure is equivalent to moving a small reflector into the position of the first zero point and then taping the distance. The reason for not calibrating the delay directly for distance is that the wire coils are not stable for very long, and the calibration would change from day to day.

How do we know the number of unit lengths? We change the frequency by a small amount and get another unit length, say 15.15 meters. Then the small distance to the first zero point will assume a new value. The difference as compared with the old 15-meter value gives the number of unit lengths. The length 15 meters is half the wavelength λ according to the formula

$$c = n\lambda$$

where c is the velocity of light and n the frequency. The precision of the present method lies in the fact that c and n are very accurately known and that they remain almost constant. For c there is the small known variation with air temperature of $10^{-6}/^{\circ}\text{C}$ and $0.4 \times 10^{-6}/\text{mm Hg}$ of pressure. For humidity the change is only $0.05 \times 10^{-6}/\text{mm Hg}$ of vapor pressure. Finally, the color dependence is $0.6 \times 10^{-6}/100 \text{ \AA}$. Phototube and optics have a comparatively narrow maximum of green-yellow at 5600 Å. The frequency n is held constant by an oscillating quartz crystal as for ordinary radio stations.

One source of error remains. The reaction

TABLE 1

Base Line		Distance, m	Velocity km/sec	Mean of Base	Weight
Swedish					
(13-15)	1	6,910	299,793.02	3.05	10
	1	6,910	3.08		
	2	5,400	3.17	2.80	8
	2	5,400	2.43		
	3	7,320	3.37	3.37	7
Australian					
(16)	1	6,440	2.64	2.44	11
	1	6,440	2.05		
	1	6,440	2.64		
	2	9,660	2.41	2.46	14
	2	9,660	2.50		
	3	6,440	3.61	3.61	3
	4	11,100	3.21	3.21	6
Total mean: 299,792.85 \pm 0.16 km/sec					

TABLE 2

Base Line		Distance, m	Velocity, km/sec	Weight
Australian (16)				
	1	6,440	299,792.50	6
English (17)				
	1	11,260	2.40	11
	2	24,830	2.20	25
U. S. A. (18)				
	1	1,380	4.06	1
	2	12,800	4.27	13
	3	3,120	2.73	3
	4	2,130	1.69	2
Weighted mean: 299,792.75 \pm 0.34 km/sec				

time of the phototube is the transit time for the photoelectrons to pass through the tube. This time is equivalent to a small light path or additional distance for the light to pass. An accidentally unsymmetrical illumination on the photocathode, however, can change that additional distance for only one of the reflector or light conductor systems, so that it will not be totally canceled by taking the difference between the two. For the first Geodimeters of model 1 this error was somewhat troublesome. If the light conductor became distorted after calibration, the error was constant. The Australians generally compensated for this error by measuring the distance first undivided and then divided into two parts. In Sweden we had two reflectors in the field and measured the distance between them; the reflectors had first

been placed close to their zero points and the light conductor was omitted. Nowada such measuring technique is unnecessary. For the Geodimeters of model 2 after 1956 the transit time errors are kept smaller than 2 cm. If errors from atmosphere, frequency, and color are added we get $2 \text{ cm} + D \times 10^{-6}$ as an expression for the accuracy of a Geodimeter determination, where D is the distance.

To produce a reliable value for the velocity of light I have combined the Australian and Swedish measurements on base lines, where the distances are known. The dispersion of the velocity values yields information directly on the corresponding accuracy in distance measurements. The results are shown in Tables 1 and 2.

The weights are proportional to the distances and the square root of the number of determinations; the determinations are from different occasions. The mean error of the mean is 1 part in 2 million. For a 10-km base and single measurement we get from the tables an uncertainty of 1 : 800,000. Inasmuch as the error of the base is included, the Geodimeter uncertainty is smaller than 10^{-8} , or the same as for a taped base line. This is for level ground. For greater distances over hilly country, Table 3, which gives results from Swedish first-order triangle sides, can be consulted.

The carrying through of a measurement requires the sun to be at least 5° below the horizon. Since the light has to go to and fro, visibility must be double the distance to be measured. If you can easily discern a 60-km-distant terrain contour in the evening, you can also usually measure a 30-km distant after sunset.

The total station weight of the large Geodimeter is 150 kg. For transport the instrument can be divided into 50-kg sections. At the station it is usually placed on one of the transport cases, or it can be placed on a low broad tripod. A circular guide for a small roller table is supplied so that the Geodimeter can easily be turned to the desired direction.

One man can carry the reflector and put it into position. As this glass-prism system need be directed only within $\pm 10^\circ$, it can be hoisted up in a tree and measurements made toward it. The tree can be steadied for definition of the position. In order to direct the Geodimeter toward the reflectors they are easily identified

TABLE 3. Geodimeter-Checked First-Order Triangle Sides

Geodimeter Point and Date	No. of Tr.*	Model†	Distance, m	Difference, ‡ cm	Relative Deviation, 1:
S I, 1947	7	0	11,031.89		
51		2	.87	+ 3	370,000
53		3	.87		
Blahöjden, May 1950	6	1	28,071.72		
October 1950		1	.75	-16	180,000
November 1950		2	.70		
Distunturi, 1949	15	1	30,918.54	+69	45,000
53		3	.49		
Neberg, 1952	1	1°	18,877.73	-57	33,000
55		3	.68		
Grundet, 1949	3	1	20,203.57	+32	63,000
en, 1950	2	1	23,506.30	+ 5	470,000
elles, 1951	?	2	12,016.62	+22	55,000
Kakavaara I, 1951	7	2	35,078.86	+ 3	1,200,000
1951	8	2	36,358.41	- 2	1,800,000
Kasberget, 1952	1	1°	12,088.40	+13	93,000
Sjöberg, 1952	9	1°	32,449.08	+51	64,000
Backen, 1952	13	1°	37,070.19	+19	200,000
Myrberg, 1952	5	1°	36,096.20	-24	150,000
Knebösklint, 1952	1	1°	25,312.92	+ 4	630,000
Emberg, 1953	6	3	32,630.46	+ 9	360,000
Berget, 1953	12	3	21,905.38	- 5	440,000
Taberg, 1953	12	3	33,183.44	+56	59,000
Myrberg (polygon), 1954	8	3	112,847.41	+132	85,000
Grundet (polygon), 1955	10	3	97,000	+32	300,000
Berget, 1956	8	3	28,008.04	- 4	700,000
Berget (polygon), 1957	4	3	83,000	+23	360,000
Åsen, 1958	17	3	37,330.72	+10	370,000
Stapuoda, 1958	9	3	31,993.60	+74	43,000
S II, 1958	4	3	8,604.76	- 7	120,000
Åla (polygon), 1958	8	3	39,388.03	+52	76,000

Number of net triangles to nearest base line.

0 = Laboratory model; 1 = experimental model; 1° = frequency uncertainty 1 : 300,000; 2 = Aga's prototype; 3 = Aga's series model.

‡ Difference is from coordinate values. Thus, old methods give an average accuracy of 1 : 100,000.

small searchlight and a field glass. The angle can also be measured by means of a theodolite.

The big Geodimeter is especially designed for first-order measurements or in general where the highest attainable precision is desirable for distances of 5 to 40 km. With care, the accuracy is of the same order as for precision taped distances. Therefore the instrument is always capable of correcting doubtful sides in a triangulation net. Among Swedish measurements, we have, for example, two sides, close to base line. In spite of this proximity the accuracy is poor. Renewed triangulation, calculation, and checks resulted in a close agreement with the

Geodimeter value. On the whole, about no Swedish Geodimeter distance can it be said that another value would be more accurate.

If the requirements for range and accuracy are lowered there will be a corresponding gain in weight, rapidity and convenience of handling the instrument. So we have model 3 of 37 kg complete station weight with tripod and power plant. For a moment, let us return to Figure 1. The delay of the electrical feeding of the phototube from the transmitter was adjusted to zero deflection of model 1, and this defined the position of the reflector. Then the delay was checked by having the light pass a built-in light conductor, the length of which was adjusted

to give zero deflection for the actual delay. For model 3 there is no variable light conductor, only a short fixed one, used to check the delay circuits as a whole. The delay dial is calibrated directly to distance. With the light through the short conductor the zero position of the delay dial at the low end is read off. Then we put in almost the total delay and reach the next zero point. The difference of the readings should be half the unit length of distance, that is 50 meters for model 3. Greater accuracy can be attained by placing the reflector close to one of these two end readings of the delay. Or the actual delay can be checked by taping the distance to a small auxiliary reflector close to the instrument.

The accuracy of model 3 is $8 \text{ cm} + 1.5D \times 10^{-6}$. With the extra precautions just mentioned it is increased to, say, $5 \text{ cm} + 1.5D \times 10^{-6}$ for distances up to 20 km.

Model 3 is very rapid in use. The Swedish water power service has had good results with this instrument. On an average, 20 minutes is required for one determination and 15 distances in one night is considered normal speed. This is for polar measurements. Model 3 seems to be utilized most efficiently for determinations from a central point. A day group distributes a series of reflectors over the area surrounding the Geodimeter point; the night group measures distances and angles. There is no possibility of false objects in these measurements. For check the instrument can be moved half a meter and the determination repeated. Visibility is usually sufficient if distances shorter than 10 km are chosen. The central points for the instrument may be the corners of a polygon, possibly obtained beforehand by some other method. Doubtful sides of the polygon can be checked by means of the Geodimeter, if necessary, after bisecting them: the instrument at mid-point and reflectors at the ends.

Provided special precautions are taken, model 3 is useful for certain first-order purposes, where the terrain is difficult for transporting model 2. Model 2 has many other special uses, for distances across water, among houses, or over terrain where other methods yield unreliable results.

For local survey measurements there is also the smallest Geodimeter, model 4, yielding al-

most centimeter accuracy up to 5 km, or approximately $1.5 \text{ cm} + 3D \times 10^{-6}$. This instrument can be operated by daylight, at least up to 500 meters. It has various uses. The photogrammetrist rapidly fixes his supporting point; the land surveyor measures with accuracy in spite of densely built-up areas or in traffic; large-scale builders are less troubled by difficult terrain, intervening water, etc. The total static weight for this instrument is 25 kg; the instrument itself weighs 17 kg. One measurement completed in 10 minutes, and another 10 minutes suffices for the calculation. Table 4 gives check results from different model 4 instruments.

It is almost impossible to give any idea of the future design of this instrument, for this depends on the development of very different details, especially light sources, modulator, and phototube. As for measurements, all taped baselines will become superfluous, as they already are in Sweden, because the connected first-order side is measured directly with considerable higher accuracy than previously. For second order, model 3 will be used in combination with theodolites or other distance instruments, as they have been in northern Sweden by the Swedish water power service. This service has covered extensive areas and has saved considerable money and time by using these modern methods. On the whole, this combined type of survey will probably be more profitable than pure trilateration, which will be kept for special cases of utmost precision. Finally, for the troublesome shorter measurements, model 4, because of its high precision, will replace the tape in urban work and in large-scale engineering projects.

There are some who think that additional determinations add strength to the survey. They should remember that each frequent measurement gives an independent distance and that, if the individual results are free from contradiction, the combined result is correct. The special function of the reflector system also excludes the possibility of a false object or signal. Moreover, by moving the Geodimeter or the reflector, or both of them, half a meter and repeating the measurement, one or several additional determinations can be obtained. Perhaps the additional determinations will be required

TABLE 4. Summary of Results Obtained at Aga with 11 Different Instruments, type NASM-4 Distance as determined with NASM-2A: $D = 3067.611 \pm 0.007$ m (one observation).

Date 1959	Instrument No.	F_1	F_2	F_3	Mean cm	Difference from Total Mean
		decimals only, cm				
Oct. 16	09	63	60	60	61.0	-1.3
Oct. 27	08	64	62	63	63.7	+1.4
March 2	10	63	62	60	61.7	-.6
March 11	11	62	64	60	62.0	-.3
March 13	12	63	63	63	63.0	+.7
March 17	14	62	62	61	61.7	-.6
March 20	15	60	61	63	61.3	-1.0
March 24	13	64	63	65	64.0	+1.3
April 1	17	65	63	61	63.0	+.7
April 14	18	63	65	63	63.7	+1.4
April 21	16	61	61	61	61.0	-1.3
April 21 (night)	16	64	59	60	61.0	-1.3
	Total	62.8 ± 0.4	62.1 ± 0.5	61.7 ± 0.5	62.3 ± 0.3	

Probable error of one observation (three frequencies) = ± 11 mm.

Average measuring time = 12 minutes.

beginning, but very soon they will be found superfluous.

For distances greater than 10 km the accuracy is limited mainly by the available meteorological information. Temperature is the most important; it is now read off at one or both ends of the distance. For increased accuracy intermediate temperatures must be known. Dr. Andersson at the Survey Office has suggested a theodolite determination of the actual elevation angle to the reflector. The elevation difference being known, the deviation from normal angle should give some idea of the total temperature influence. A rough estimate leads to the following formula:

$$\Delta = \frac{\cos(\beta + \epsilon)}{\cos \beta} - 1$$

Multiplying the distance by the factor Δ gives the length correction to be applied. ϵ is the elevation angle from that for air of even temperature, observed by the aid of a theodolite

and positive downward. β , always positive, is the inclination of the isotherms at the geodimeter end and in the direction of the distant reflector. β follows the inclination α of the ground; it is proportionally smaller for big α , and certainly smaller the stronger the wind. An approximate average may be $\beta = \alpha/2$. If we inspect data of altitude measurements for moderate country with inclinations smaller than $1/10$, Δ only exceptionally exceeds 10^{-6} . No practical check of the formula has been made.

Theoretically we can determine the actual integral of atmospheric influence by measuring the distance with two different colors of the light source, e.g. blue and yellow. They differ by 3 per cent of the total atmospheric influence. Knowing this value from the distance reading, we also know the total amount, which is 30 times greater. A relative accuracy of better than 0.5 mm of distance for 10 km, is sought. The main obstacle is the thermoturbulence of the atmosphere.

Use of the Geodimeter by the Coast and Geodetic Survey

Abstract

PAUL D. THOMAS

*U. S. Coast and Geodetic Survey
Washington, D. C.*

The Coast and Geodetic Survey has measured 84 geodetic lines with Geodimeters, models 1 and 2, over a 5-year period. On the basis of this experience, a routine of field operations has been developed and a manual of operation is being prepared. Maintenance problems have been negligible, and only minor modifications of the in-

struments have been made. The Coast and Geodetic Survey recommends the Geodimeter (model 1 or 2), for measuring first-order triangulation base lines, trilateration, and traverse and for calibrating electronic distance-measuring equipment such as the Tellurometer.

The Model 3 Geodimeter for the Extension of Control for California Highways

Abstract

JAMES D. CARTER

*Division of Highways
Sacramento, California*

The California Division of Highways purchased a model 3 Geodimeter in August 1957; its experience to date consists of 3000 hours of observing time, or well in excess of 15,000 miles of line. The Geodimeter crew is equipped with four-wheel-drive vehicles, each with 65-watt radios, plus several small handset radios for walk-in reflector stations.

The range of this instrument may be limited to distances as short as 1 mile in a dense smoke haze, increasing rapidly with better visibility to an outer limit of 25 to 30 miles in the clear air of higher elevations.

The standard procedure consists of making the normal set of distance readings recommended in the Geodimeter instruction manual. The time required for a set of readings should be about 45 minutes per line to yield design accuracy. If the Geodimeter is calibrated at 300-hour intervals, it will contain no systematic errors and will not exceed the following accidental errors for a series of measurements: 0 to ± 0.15 ft, 25 per cent; ± 0.15 to ± 0.25 ft, 50 per cent; and ± 0.25 to ± 0.35 ft for the remaining 25 per cent of the distances.

The model 3 is also capable of geodetic accuracy provided a technique known as short-base calibration is employed. The first measuring frequency, giving the last four digits of the distance, is the only frequency that needs to be tested. A normal set of readings is made on the

desired line, and the first two columns of calibration readings are eliminated. The last four digits of this distance are then computed, e.g. 25.75 meters. A small plastic reflector is then placed at this distance from the Geodimeter and the distance carefully measured by tape and the instrument. If the correction resulting from the comparison is applied to the main line, approximately 80 to 90 per cent of the instrumental error will have been removed. Where the economies of speed are more important than the need for high accuracy, a third method consists in reducing the usual set of distance readings by 40 per cent and speeding up the remaining readings. This increases the range of accidental errors to an average of ± 0.50 ft.

Errors quoted are for the Geodimeter only; other errors, which usually exceed these, must be taken into account in planning the survey. Sources of such errors may be (1) the vertical angles, (2) the horizontal angles, and (3) the quality of the control ties.

In addition, the accidental errors of these instruments tend to cancel through a series of courses and may give a small closing error, which, however, is of little value in estimating the accuracy of any one course in that traverse.

Our work indicated a time reduction of approximately 30 per cent in the areas suitable to tape-traversing to at least 200 per cent in the mountains.

An Electro-Optical Device for Measuring Distance

ARNE BJERHAMMAR,

*Tekniska Hogskolan,
Stockholm 70, Sweden*

In geodesy and surveying accurate determinations of distances are of great importance. During World War II many new principles for electronic distance measuring were invented, and some of them have been applied to geodesy. When the greatest geodetic accuracy is needed, however, the electronic methods have, up till now, not always been successful. One of the main disadvantages of purely electronic methods is that ground reflections affect the final results. Furthermore, ground conductivity seems to introduce special complications for longer wavelengths. Finally, meteorological conditions are of great importance for a determination of the velocity of propagation.

Some of the difficulties involved in electronic methods can be eliminated by using light as a carrier instead of electronic waves. This method of electro-optical measurement was first applied for purely scientific investigations. Thus the classical determinations of the velocity of light were made by means of devices that could have been used for distance measuring if only the velocity of light had been accurately determined (Fizeau, Foucault, Michelson, Karolus, Mittelschmidt, Hüttel, and Anderson). These determinations of the velocity of light were perhaps considered of no great practical importance at the time. It is evident, however, that Michelson suggested the application of his method of determination of distances for geodetic purposes.

The first patent of a real electro-optical distance-measuring device was applied for by Wolff in 1939. He made use of an ultrasonic light relay to modulate the light. Later Bergstrand and Hanson designed electro-optical devices for distance measuring with geodetic applications. Measurements made by Bergstrand have shown that the electro-optical distance-measuring methods can give an accuracy comparable with the greatest geodetic accuracy.

One of the main limitations of the electro-optical methods used up till now has been the complexity of the devices. At the Division of Geodesy of the Royal Institute of Technology at Stockholm a research program for investigating new ways of making electro-optical distance measurements was started in 1951. As a result of these investigations one new design for Kerr cell-operated light modulators, another for concentrated arc lamps, and a third for quartz crystal modulation have been developed.

A short description will be given below of the general principles of light modulation by means of a quartz crystal, and also the special design developed for distance measuring (here called 'terrameter') will be briefly discussed.

QUARTZ CRYSTAL AS A LIGHT MODULATOR

The quartz crystal is a 'uniaxial' crystal, that is, the crystal has one optical axis. On the optical axis, when light passes through the crystal parallel to the optical axis, the light will be uniaxially refracted. In all other directions the light will be doubly refracted. It is easy to demonstrate this for polarized light passing through the crystal. There will be a rotation of the polarization plane. Not all quartz crystals rotate the polarization plane in the same direction, and therefore we speak of left- and right-handed crystals. Exposure of the quartz crystal to mechanical strain will result in a piezoelectric effect: there will be a static electrical voltage directly correlated to the strength of the strain. Furthermore, the optical properties of the crystal will be changed. The crystal will be doubly refracting in all directions. If the strain is taken away, the crystal will again act as a uniaxial crystal. Finally, the quartz crystal is an outstanding device for frequency control. It will resonate with a frequency that is defined ve-

cisely from the cut of the crystal and from geometrical dimensions. The quality of a crystal can be measured by the time the crystal continues to vibrate in the resonance frequency without any generating power. For practical applications this quality measure can be reduced by the Q value, defined as the relation between the electric reactance and resistance of the crystal.

The electrical equivalence of a quartz crystal tuned circuit, of which there are two types: series-resonance circuit, and the parallel resonance circuit; and the quartz crystal can work in both modes. The series-resonance circuit is a single unclosed circuit including a capacitance and an inductance. For the resonating frequency the impedance of the capacitance is exactly equal to the impedance of the inductance, and because the two impedances compensate each other the final impedance will be equal to zero. In all practical circuits there will be some additional resistance. For a good quartz crystal (8 Mc/s) this resistance is only about 50 ohms. A quartz crystal that is operated in series resonance is capable of filtering out practically all but the proper resonance frequency. By means of this filtering effect it is possible to stabilize the frequency of an external generator. However, the crystal itself cannot work very well as a self-oscillating generator if it is operated in series resonance. When a crystal is used as an oscillator, therefore, it is normally operated in parallel resonance.

The corresponding equivalent circuit is a parallel circuit including a capacitor, an inductance, and some additional resistance. The impedance of the capacitance and the inductance must be equal, but the total impedance of the parallel circuit is very high here. For a good crystal (8 Mc/s) the total impedance can be about 1,000 ohms. When the crystal is operated in parallel mode it easily forms a self-oscillating circuit.

Evidently, every crystal has two working frequencies, the resonant and the antiresonant frequency. When the crystal is operated at either of these two frequencies it is extremely sensitive. Thus the necessary power to be applied for maximum activity is normally a fraction of a watt. The mechanical activity of the crystal is a function of its electrical charge, and the

whole phenomenon can be regarded as an inverse piezoelectric effect.

If a crystal is to be applied as an electro-optical light modulator different ways of modulation can be considered. One is to use the diffraction effects caused by ultrasonic waves in the crystal. The method we have found most suitable for our applications is to utilize the double-refracting effect. A practical design for this type of operation consists of two crossed polarizers with the crystal between them. If the light is passing parallel to the optical axis of the crystal, there will be almost no light after the last polarizer. If the crystal is oscillating, light will pass in bundles of a frequency twice as high as the oscillator frequency. If the last polarizer is turned 45° in any direction the light will be modulated with the same frequency as the applied electrical frequency. It can easily be proved mathematically that the two signals obtained from the last polarizer oriented at $+45^\circ$ and -45° are phase inverted in relation to each other. This interesting optical relation has been applied to the design of the electro-optical distance meter.

Light modulation by means of crystals seems to be more favorable for distance measuring than Kerr-cell modulation. The main difference is that the distortion is much smaller for the crystal than for the Kerr cell. Especially if a high modulation degree is used, the Kerr cell will give a considerable distortion. Furthermore, the crystal requires only a fraction of the power needed for Kerr-cell operation at the same frequency. On the other hand, the Kerr cell has the advantage of being adaptable to any arbitrary frequency.

DEFINITION OF MEASURING POINTS FOR ELECTRO-OPTICAL DISTANCE MEASURING

If light is modulated by means of a crystal, the crystal must be operated at a frequency rather close to the resonant frequency. Only under such circumstances will there be a favorable relation between the power dissipation and the electro-optical effect. The electro-optical light modulation is therefore limited to a number of frequencies that are kept constant during the whole measuring operation. With respect to the wave form of the modulated light the possibilities are also limited. It is evident that

only wave forms of sinusoidal type can be generated by the crystal oscillator. Even if the wave form will never be a pure sine wave, it can at least be interpreted as a sine wave with some additional distortion. In order to define a measuring point by the aid of sine waves one of the following methods can be applied.

(a) *Additive mixing.* If the measuring points are to be located optically a fraction of the outgoing light bundle can be added to the returning light bundle. Then the sum of the two light bundles will be more or less modulated. If the two sine waves are in phase there will be maximum modulation; if they are out of phase, minimum modulation. Finally, if the two sine waves are equal in amplitude there will be no modulation at all at the minimum point.

Thus we have:

$$\frac{\begin{array}{l} \text{First sine wave } \sin \omega t \\ \text{Second sine wave } \sin (\omega t + \Delta) \end{array}}{\text{Sum } 2 \sin [\omega t + (\Delta/2)] \cdot \cos (\Delta/2)}$$

where

ω = angular frequency.

Δ = phase difference between the two waves.

t = time.

Normally, it is not possible to detect the modulation with the naked eye. Perhaps the simplest way of detecting it is to add a phototube and an electronic voltmeter device. The voltmeter will then give the absolute value of the signals, and we obtain

$$e_0 = |c \cdot \cos (\Delta/2)|$$

and $e_0 = 0$ if $\Delta = 180^\circ$.

The minimum points are here uniquely defined, and the definition is very sharp.

After addition of sine waves of different amplitude the corresponding final voltmeter reading will be

$$e_1 = |c \cdot \cos (\Delta/2)| + k$$

In this case the minimum current never goes down to zero and the definition of the measuring point will be less sharp (at least, if the relative variation of voltage is considered).

(b) *Multiplicative mixing.* An ordinary elec-

tronic multigrid tube can be used as a device for multiplicative mixing. With a suitable tube there will be a multiplication of signals that are fed to two different control grids. Such a tube is useful as a phase-sensitive device, and the following measuring method has been tested.

The optical signal from a light-modulating device is converted into an electronic signal in a phototube. The new signal is then to be compared in phase with a primary signal of the same frequency. To make a phase comparison possible we split the primary signal into two components so that we also obtain an additional signal one-half wavelength out of phase. For the phase measurement we use two multigrid tubes and feed the 'optical signal' to the first grids. The primary signal is fed to the suppressor grid of the first tube, and the phase inverted primary signal to the corresponding grid of the second tube. The current through the two tubes will then be given by

$$i_1 = \frac{1}{2\pi} \int_{t=0}^{t=2\pi/\omega} k \sin \omega t \cdot \sin (\omega t + \Delta) dt$$

and

$$i_2 = \frac{1}{2\pi} \int_{t=0}^{t=2\pi/\omega} k \sin (\omega t + \pi) \cdot \sin (\omega t + \Delta) dt$$

where

ω = frequency in radians ($f \cdot 2\pi$).

t = time.

i = current.

From this we obtain the difference between the two currents:

$$i_1 - i_2 = k \cdot \cos \Delta$$

and

$$i_1 - i_2 = 0 \text{ for } \Delta = 90^\circ \text{ and } \Delta = 270^\circ$$

This method of measuring has been tested with good results. Its disadvantage is that it is hard to find tubes sufficiently similar for this type of operation. Therefore, a modified design with only one tube was tested. In the final design additive mixing has been used for the definition of the measuring points.

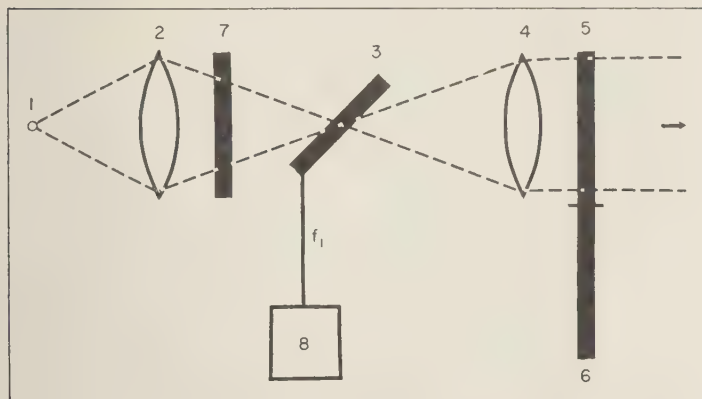


Fig. 1. The optical system of the Terrameter.

- | | |
|------------------|---------------------|
| 1, light source. | 3, optical crystal. |
| 2, 4, lens. | 5, 6, 7, polarizer. |
| 8, oscillator. | |

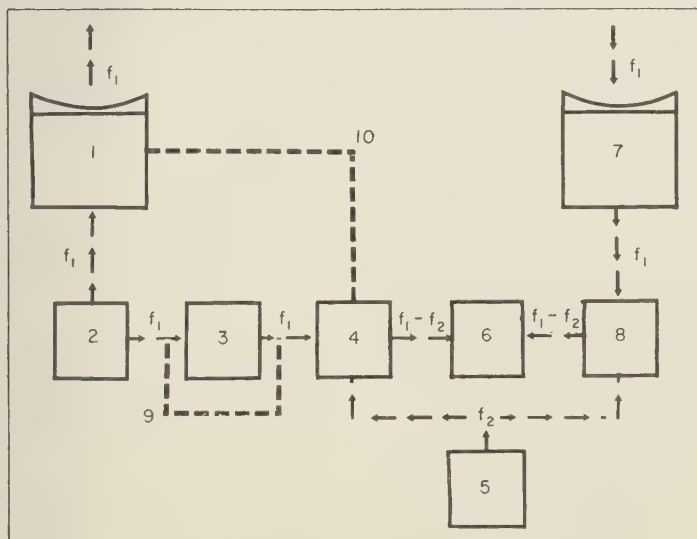


Fig. 2. Block diagram of the Terrameter.

- | | |
|--------------------------|---|
| 1, light modulator. | 6, phase indicator (or zero instrument). |
| 2, oscillator (f_1). | 7, optical receiver. |
| 3, delayer. | 8, photo multiplier. |
| 4, mixer. | 9, alternative by-passing. |
| 5, oscillator (f_s). | 10, alternative modulation of frequency $f_1 - f_s$. |

THE ELECTRONIC DESIGN OF THE TERRAMETER

The terrameter includes two main parts, the transmitter and the receiver, and a mirror is used to reflect the light signal back to the instrument.

The transmitter is shown in Figure 1. Here the light source emits light which after polarization is collected in a lens and finally concentrated in the center of the crystal. The outgoing light from the crystal is then collected in a new lens and is made parallel. In the front of this lens there is a polarizer with two polarization axes perpendicular to each other. When a measurement is made the last polarizer is operated so that readings are made by using both polarization planes.

A block diagram of the whole assembly is given in Figure 2. The optical signals of frequency f_1 from the crystal, 2, are reflected by a distant mirror and then finally received in an optical system of the receiver and projected on the cathode of a photomultiplier, 8. This multiplier is also fed with an auxiliary frequency f_2 from a crystal oscillator, 5. The frequencies f_1 and f_2 are mixed in a multigrid tube, 4, and the beat frequency $f_1 - f_2$ is fed to a grid of another multigrid tube, 6. This grid is also fed with corresponding beat frequency from the photomultiplier, 8. In this multigrid tube there will be a multiplicative mixing of the signals, and therefore the anode current will vary as a function of the phase relation between the two

beat signals. These two beat signals represent the same phase relation as that of the outgoing and incoming signals.

A mathematical explanation of the function of the instrument can be given in the following way: When the last polarizer is oriented $+45^\circ$ we obtain the signal

$$s_e' \simeq \sin \omega t$$

and when the polarizer is oriented -45°

$$s_e'' \simeq \sin (\omega t + \pi)$$

The optical signal is reflected by the mirror and finally received by the photomultiplier. The signal has then received phase information and can be represented by the formula

$$s_e \simeq \sin [\omega t - (2\pi d/\lambda)]$$

Multiplicative mixing with the auxiliary signal $s_1 \simeq \sin \omega_1 t$ will give

$$\begin{aligned} s_2 &\simeq \sin \left(\omega t - \frac{2\pi d}{\lambda} \right) \sin \omega_1 t \\ &= \frac{1}{2} \cos \left\{ (\omega - \omega_1)t - \frac{2\pi d}{\lambda} \right\} \\ &\quad - \frac{1}{2} \cos \left\{ (\omega + \omega_1)t - \frac{2\pi d}{\lambda} \right\} \end{aligned}$$

where

- d = optical length.
- t = time.
- ω, ω_1 = angular frequency.
- λ = wavelength at frequency f .

We have now split the original signal into two components, the first one including the angular frequency $(\omega - \omega_1)$ and the second one $(\omega + \omega_1)$. The last part of the information is filtered out, and we obtain

$$s_3 \simeq \cos [(\omega - \omega_1)t - (2\pi d/\lambda)]$$

This signal includes the same phase information as the original signal. The important difference is that the new signal is a low-frequency one and therefore much more suitable for distance measuring.

An example will demonstrate this. If

$$f = 10,000,000 \text{ p/s}$$

$$f_1 = 9,999,900 \text{ p/s}$$

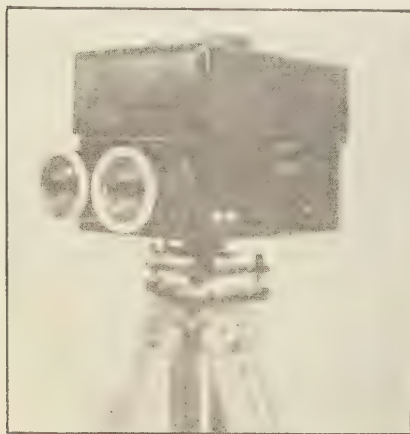


Fig. 3. A Terrameter.

$$f - f_1 = 100 \text{ p/s}$$

Thus, if the optical path is changed by Δd meters, the corresponding electronic change will

$$\Delta d \cdot f / (f_1 - f_2) = 100,000 \text{ meters}$$

This method of 'double superheterodyning' has been applied to all electro-optical distance-measuring instruments developed at the Division of Geodesy in Stockholm. It has also been used as a purely electronic measuring method with very satisfactory results.

SPECIFICATIONS OF THE 'TERRAMETER'

Lens diameters, 80 mm
 Optical length, 250 mm
 Weight of the instrument, ~ 9 kg
 Lamp for light modulation, ~ 0.7 watt
 Power for transmitter, ~ 2.0 watts
 Power for receiver, ~ 2.0 watts
 Power for heater, 0.8 watts
 Total power consumption, 5.5 watts
 Working frequency, 4,000,000 p/s
 Beat frequency, 3000 p/s
 Frequency stabilizer, crystal cut AT
 Optical crystal, cut BT
 Auxiliary frequency, crystal cut AT
 Measuring tolerances:
 Internal reading off errors for good optical signals, ± 1 cm
 Layer: 'helipot'
 This commercial delaylayer has systematic errors up to 50 cm; after calibration, ± 2 to 3 cm
 Frequency errors (without thermostat) $\sim 10^{-5}$
 Maximum distances:
 For practical applications, 3 to 5 km
 Shortest distance:
 For practical applications, 50 m

The statistical analysis of the author's experimental data, included in the report presented to the meeting, is omitted from this condensation. We give instead, in the following, the author's own abstract of the investigation. *Ed.*]

ANALYSIS OF THE LIGHT MODULATOR OF THE TERRAMETER

A theoretical analysis of the modulated light from a quartz BT crystal has shown that, at a

proper adjustment of the polarizers, the fundamental frequency is obtained without any distortion of even harmonics. As odd harmonics give no errors in a suitable phase discriminator, this type of light modulator is well adapted for accurate distance measurements. A similar investigation of the modulated light from a Kerr cell shows that even harmonics dominate completely for higher modulation percentage. An investigation of the modulation phase in different parts of a convergent light bundle from an oscillating BT crystal has shown that at larger aperture serious phase distortion can be found. These errors have been eliminated by focusing the light, not to the crystal itself, but to a point somewhat behind the crystal. Such errors can be diminished by using a smaller aperture. Studies of the modulation percentage at different orientation have shown that it is possible to obtain as much as 70 per cent modulation. This part of the investigation is verified by complete photographic recordings.

MEASUREMENTS WITH THE TERRAMETER

Measurements made with the instrument have been checked for accidental and systematic errors by statistical methods. The systematic errors of an electromagnetic delay line have been studied by Fourier analysis. The zero point drift of the instrument has been studied by a distribution for the ratio between the variance and the mean square successive difference. Fisher's distribution has been used for the study of the variance ratio for different measurements. With the aid of correlation analysis the measurements from different groups have been studied and compared. It has been found that at a distance of 1 km the standard deviation of a single measurement was ± 11 mm (10 measurements).

This investigation was made at the Division of Geodesy of the Royal Institute of Technology, Stockholm. Collaborators were Messrs. B. O. Malmberg, G. Zelinger, A. Forsberg, and E. Bruchfeld. Statens Tekniska Forskningsråd sponsored the investigations.

The 'Tellurometer' System—New Applications to Geodesy and Hydrography

R. D. SMITH

*Tellurometer (Pty), Ltd
Cape Town, Union of South Africa*

INTRODUCTION

The Tellurometer system, in the form of Microdistancer equipment for overland measurements, was introduced early in 1957. It may be fairly stated that since that date a new vista has been opened to the surveyor—the use of accurate radio line measurements in the art of surveying. As we all know, this has already had a profound effect upon land survey practices. It is the purpose of this paper to discuss new developments of the system as they may be applied to geodesy or hydrography.

Principles of the Tellurometer system. The general theoretical considerations of the Tellurometer system of distance measurement are adequately covered in the references at the end of this paper.

The Tellurometer in conventional geodetic application. The short history of events during the past two years has served to establish the direction that developmental work must follow to meet the varying requirements of the land surveyor.

Its actual use, therefore, in the fields of geodetic, topographical, engineering, and military survey has furnished the information upon which instrument design is now to be based. This statement refers to the Tellurometer system (Microdistancer equipment) as used to measure distances between static land points where the limitation of range is that imposed by the normal maximum line of sight obtainable over average terrain, which is probably of the order of 35 miles.

Requirements for future Microdistancer equipment are:

(a) Reduction in weight. Transistorization to be utilized efficiently to the full, to maintain the basic requirement of easy portability and ability to withstand rough treatment.

(b) Reduction in power drain, and the evaluation of all battery improvements to enable full advantage to be taken of (a).

(c) Combination of function of master and remote station.

(d) Greater instrumental accuracy with methods of field-checking reference standards.

(e) Development of 'nonswinging' Tellurometer system.

These advances are all being actively pursued and improvements are being steadily introduced into new equipment. The dual master remote function has already been incorporated in a new military version of the Tellurometer system. It is not, however, a function of this paper to discuss the geodetic Tellurometer system, but rather the new applications to which the basic system can be put.

2. NEW APPLICATIONS

The new applications now contemplated are: (a) Those relating to the positioning of surface vessels as in hydrography; this subject is dealt with in section 4. (b) Those relating to the measurement of long lines, or to the positioning of distant points with the aid of an aircraft, an intermediate station, and to possibilities of positioning the aircraft itself for certain purposes, e.g. photogrammetry. This subject is dealt with in section 5.

Before discussing these subjects it is necessary to give consideration to some of the problems primarily occasioned by movement and involving range and accuracy.

3. THEORETICAL CONSIDERATIONS: MOVEMENT, RANGE, AND ACCURACY

Movement

As has been said, the present Microdistancer version of the Tellurometer system was d

ned to measure distances between two static points on land. As soon as its use on surface vessels or aircraft is contemplated, the question of movement makes the deliberate character of the operational procedure that has heretofore been adopted either difficult or impossible. In the present discussion, however, movement has been dealt with as a complication affecting other problems, namely range, accuracy, and presentation methods, and is referred to under these heads.

Range

Under the heading of range, factors concerning carrier frequency, carrier power, and antenna beamwidth must be considered together, as they are closely interrelated.

Carrier frequency. The carrier frequencies of the geodetic Tellurometer system (Microstancer) are in the 10-cm band.

It can be shown that greater ranges can be achieved with longer wavelengths. Frequencies higher than 10 cm will obviously enable narrower beamwidths to be achieved with comparatively small reflector systems, and the narrower beams utilized may tend to overcome absorption losses of the atmosphere under these higher frequency conditions.

Transmitter power. The transmitter now utilized provides an output of approximately 100 milliwatts. At lower carrier frequencies it will be possible to obtain higher power outputs at probably greater efficiencies. At frequencies higher than 10 cm the efficiency of klystrons will definitely fall off considerably and less power will be available. Down to 3 cm efficient power outputs of up to 100 milliwatts should be available for transmission purposes. Below 3 cm still less power may be expected.

Beamwidths. At 10 cm with 18-inch reflectors, it is possible to obtain beamwidths of the order of 20°. At lower frequencies larger reflectors will be necessary to maintain beamwidths of this value or considerably larger if narrower beams are required. At higher frequencies smaller beamwidths become more easily obtainable with small-diameter dishes. The selection of beamwidths that are narrow in one particular plane and broad in another can also be considered, and for certain applications of relatively short range it may be possible to have

comparatively nondirectional transmission characteristics.

Accuracy

In dealing with accuracy many interrelated factors must be considered.

Modulation pattern frequencies. Modulation pattern frequencies of 10 Mc/s are utilized in the geodetic version of the Tellurometer system. Other applications either may not require a high degree of accuracy or, by virtue of the fact that moving platforms are involved, may require a reduction in the pattern frequency. Lower modulation frequencies, providing a coarser scale but more susceptible to reflection problems and larger errors, may be introduced under certain conditions.

Ambiguity resolutions. Ambiguity resolutions during the usual process of deliberate measurement can be easily achieved by a manual switching process. Under conditions of movement, however, they have to be dealt with somewhat differently. For slow-moving targets, rapid switching controlled by the master equipment will permit the ambiguity resolution to be deduced provided that remote switching is kept in step with the master unit. This system will be satisfactory on the basis of manual recording, and with reduced modulation frequencies, up to speeds of 20 knots. At higher speeds it is necessary for ambiguity resolutions to be achieved on a continuous recording chart system whereby the switching processes are automatically controlled from the master and the remote switches in sequence through the necessary ambiguity resolution processes. Automatic timing permits the interrogation for ambiguity resolution to be repeated at regular, adjustable time intervals. This process should be satisfactory to speeds in excess of 100 knots. It can be readily appreciated that once movement is introduced into the system automatic pattern switching must be considered essential.

Phase comparison and presentation methods. The original cathode-ray-tube phase comparison presentation has been very satisfactory for all normal static measurements. Once movement is introduced into the system, however, alternative methods of phase presentation and recording must be considered. For slow movement, a method of manual following, by a

cursor, of the rotating phase pattern display can be very satisfactory with reduced pattern frequencies. A simple method of setting up a counter attached to this cursor will enable an operator to follow a steady or varying movement quite easily, each rotation of the cursor adding or subtracting from the original coarse reading deduced during the initial switching process. Automatic switching controlled by the master makes it easy to check back on the 10 times digit to make sure that no lane has been missed. For more elaborate systems and higher speeds, a simple follow-the-counter system is not suitable, and these ranges and phase changes must be recorded automatically. For this purpose it is necessary to make a change, and two sine waves are utilized. Phase comparison under these conditions is made by means of a synchro resolver and selsyn operating in a closed loop system. The drive system is geared up to a recording counter, which can give direct range readings in meters if a suitable modulation frequency is selected.

For the above an average refractive index must be assumed. The value 1.000325 is regarded as satisfactory for sea-level conditions. A permanent record of range is maintained by means of a gating system which records these phase changes on a continuous strip chart. The increases and decreases are linear with respect to phase difference. An electronic ambiguity resolver is incorporated in the system, to resolve an ambiguity, inherent in the particular display system used, between the 0-180° and the 180°-360° sectors. Ambiguity resolutions of the main patterns may be displayed on the same recording on a variable time basis. This distance recording can, of course, be phased into other operations, such as echo sounding and height recording, and can obviously be repeated for additional systems.

Instrumental errors. On standard instruments, under static conditions, the instrumental zero error is taken out by reversing the sense of the readings, and by meaning the A+ and A- pattern frequency readings. For slow-moving equipment it is possible for the zero error to be checked manually by switching over from A+ to A- during the setting-up procedure. With faster-moving equipment, however, this correction must be incorporated in the auto-

matic switching cycle. This can be arranged and is shown on the recording, as a mean of two readings.

External errors. Errors introduced by external means must be primarily considered under (a) ground reflection and (b) meteorological problems.

Possible ground-reflection errors are governed by the modulation frequency, the beamwidth of transmission, and the ground characteristics. With broader beams for transmission purposes the possibility of ground reflections becomes more severe; with narrower beams improvement can be expected in reflection problems. Errors introduced by reflections at lower modulation frequencies will tend to be larger under certain conditions of high reflection. These errors can be overcome, either by careful selection of sites or by making suitable aircraft flight plans.

There appears to be considerable scope for research work in the study of the meteorological conditions as they affect electronic distance measuring processes. The very fact that the present Tellurometer system utilizes the meteorological conditions at each end of the line being measured, under long line conditions, leaves the system open to possible meteorological errors. At short ranges an average meteorological condition for lower accuracy can be taken as a mean figure and a mean refractive index deduced. For long-range work utilizing aircraft the advantage of being able to record barometric and meteorological conditions in the aircraft as well as at the two or three ground points may compensate for some of the difficulties due to the greater distances between ground stations. For geodetic measurement purposes the main limitations still appear to be the wet and dry-bulb readings taken to obtain the vapor content of the air. Considerable differences of opinion exist between users as to the best method of obtaining these readings.

4. HYDROGRAPHIC APPLICATIONS

General

Hydrodist is a Tellurometer system of line measurement to obtain positional determinations of surface vessels as required in hydrographic survey. It was developed to the specification of the South African Navy Hydrographer, as

ult of requests received from various hydrographic and similar agencies for a design of this instrument, both for use in hydrography (echo-sounding procedures) and in the location of vessels engaged for dredging rivers, estuaries, or coastal waters.

The sextant-angle method, normally employed in order to obtain a fix for the moving vessel, has certain disadvantages, the main one being that work normally must be confined to periods of daylight with good visibility. A radio method of obtaining such a fix would enable work of equal, or possibly greater, accuracy to be continued regardless of weather or visibility, permitting, if necessary, round-the-clock operations for either survey craft or dredging vessels.

The system enables a fix for a moving vessel to be obtained as a function of two ranges with a high degree of accuracy. The master station is located on the moving vessel, and the remote stations at points of known control ashore.

Hydrodist Theoretical Considerations

A number of modifications to the geodetic instrument (Microdistancer) of Tellurometer equipment have been made. One modification permits a system of range presentation to be calculated directly in decimal units of distance (namely meters) instead of decimal units of miles. Calculations for meteorological conditions, which introduce negligible errors, can be eliminated. The finest reading scale comprises 100 divisions of 1 meter each, which is coarse enough to allow the operator to follow phase calculations conveniently with the ship moving at knots.

An average value for refractive index of 1.00325 has been adopted. Utilizing the accepted figure for velocity of propagation in free space, modulation frequencies are employed as follows:

	Master		Remote
A	1.498470 Mc/s	A—	1.499470 Mc/s
C	1.483485 Mc/s	A+	1.497470 Mc/s
D	1.348623 Mc/s	C	1.482485 Mc/s
		D	1.347623 Mc/s

Equipment Specification Hydrodist Model MRB2

System components. A complete system for determining the accurate position of a moving vessel comprises two single line-measuring systems. The two master instruments are located in the vessel, and the two remotes at points of known control ashore.

Each master works continuously with its own remote. An uninterrupted line of sight is necessary between each master and its remote. Maintenance of this condition is largely a function of the correct location of the remote stations ashore to meet the requirements of the coverage and range.

Equipment in the vessel. The master instruments are physically similar to the standard Microdistancer equipment, and equally portable. They may be located either within the cabin or on the deck or bridge of the vessel.

A 12/24 volt battery power supply is standard, but provision may be made for other power supplies as specified for a particular vessel.

The dipole and standard reflector are detachable from the instrument and may be used in conjunction with the double aerial mast, which is available as an accessory. That is to say, the instrument may be used either with the aerial assembly secured to the instrument itself or with the aerial detached and secured to a mast mounted above the wheelhouse on the cabin roof.

Separation of the aerial from the instrument up to maximum distances of 10 ft is contemplated; beyond this distance effective range will be lost. Coaxial cables with necessary plugs and sockets are provided for aerial separation.

The single control panel contains all operating controls, including a manually operated cursor-counter mechanism which, once set at an initial range reading, automatically indicates all changes in range due to the movement of the vessel. The range is read directly in meters.

Equipment at shore stations. The two remote instruments ashore are physically similar to present Tellurometer MRA2/MV instruments, the main differences relating to provision of separable dipole and reflector, single-panel operation, and the inclusion of an automatic pattern-following device controlled by the mas-

ter unit, for facilitating quick ambiguity resolution. Once communication has been established with the master, therefore, the remote unit can remain unattended, except for periodic checks.

Both master and remote units are provided with duplex radio telephone link and a calling device is included in the remote to attract the attention of the remote attendant.

Power supply from a 12-volt battery is normal, but where units are permanently installed or must be left in operation on a semipermanently unattended basis, power supply units to operate from mains supply are available.

Deployment of Remote Stations

Hydrodist Model MRB2 remote instruments are as portable as standard Tellurometer equipment (Microdistancer Model MRA2/MV); they can be landed from small boats, or carried by vehicles or as back packs, to the required locations.

Sites ashore must be chosen with due regard to the over-water coverage needed and the maintenance of a suitable geometric configuration for meeting the positional accuracies required. In general, the considerations are similar to those applying to a determination of the vessel's position by the sextant-angle method.

To ensure optimum accuracy of line measurements, the remote units should be so sited as to give a reasonably shallow ray-path clearance. Remotes should, therefore, preferably have a low elevation for short-range work and a higher elevation for longer ranges, subject at all times to the requirement of a clear line of sight. In practice little difficulty in the matter of siting is likely to be experienced.

Range

Hydrodist equipment is designed to give the maximum range likely to be required for coastal hydrographic survey, with line of sight from available points on shore the limiting factor.

Satisfactory operation up to distances of 20 to 25 miles may be expected when using the standard instrument reflector (giving a 20° beamwidth) on both master and remote units. Special reflectors of larger aperture will decrease this beamwidth and substantially extend the range.

Various alternative reflectors are available,

such as wide-beam reflectors for short-range work (to avoid repeated movement of aerials to give precise directivity) and specially designed reflectors (for remote units only) to provide a narrow beam in the vertical and a wide one in the horizontal plane.

All reflectors on both master and remote are provided with quick-release mechanisms to permit rapid exchange.

Accuracy

The $A + A -$ difference pattern is provided in order to free the scale reading of the A pattern from zero errors due to minor phase defects and maladjusted scale. This facility is available for checking purposes in setting up the instrument.

The adoption of the A pattern frequencies tabulated above gives rise to an instrumental error applicable to each line measurement which may be as much as 0.5 meter, regardless of the distance being measured.

Variations of actual refractive index from the mean value adopted will cause errors which under normal conditions are not likely to amount to more than about 1 part in 100,000. In extreme cases it may be as much as 1 part in 30,000, i.e. approximately 1 meter for normal maximum range measurements.

The cursor following mechanism is provided with a graduated dial of 100 divisions having a value of 1 meter per division. Reading is possible to one-quarter of a division, but for normal operation on a moving vessel 1 meter is perhaps the practical reading accuracy.

Practical tests indicate the probability that line measurements from a moving vessel under most adverse conditions will be accurate to within 1½ meters, the resultant accuracy of positional determination being dependent in measure on the trilateral configuration. In practice, roll or pitch of the vessel does, of course, affect the useful application of this accuracy.

The A pattern minus the D pattern gives the hundreds of meters up to 1000 meters. The A pattern minus the C pattern gives the thousands of meters up to 10,000 meters. No ambiguity resolution above 10,000 meters is provided, as it is regarded as certain that the vessel's position will be known to at least the nearest 10,000 meters (approximately 7 miles).

The automatic pattern-following device on the remote facilitates a quick check of the — C and A — D patterns while in motion. If equipment is employed ashore (for instance, for the purpose of extending geodetic control in the establishment of coastal survey points), the probable errors inherent in the selection of a mean value for refractive index may be avoided by taking meteorological observations in the normal manner for Tellurometer measurements and applying the appropriate correction. The resultant accuracy of an individual line measurement is almost wholly a function of the instrument error plus the scale reading accuracy. After a number of fine readings have been taken and averaged it is probable that a measurement accurate within 0.5 meter will be achieved.

Operation

For each master instrument on the vessel, an operator is required who, by means of the radio telephone link with the remote, controls the setting-up and operating procedure to be followed.

Once communication is established, the master operator proceeds to resolve ambiguities and set up the initial reading on the digital readout counter. He also checks the zero error if necessary. Provided that the vessel is proceeding at a slow speed this procedure can be followed without stopping the vessel.

The master operator, using the control handle of the cursor, then proceeds to follow the A pattern display, and the pattern rotations (as multiples of 100 meters) are automatically added to or subtracted from the readout counter.

Digital readout of range from each instrument is at all times available to the station plotter as required, and plotting may conveniently be effected by two range arms or by transparent chart overlay.

Rotation of the aerials when mounted on the double mast assembly is manually undertaken by movement of a hand wheel, and they can be directed either by visual observation or by compass reference.

Remote stations are designed for either manual or unattended operation, an important requirement of the remotes being that directiv-

ity of aerial beams should be maintained, or suitable reflectors provided, so as to ensure coverage over the area to be worked.

Frequency difference. The two masters on the vessel (with their attendant remotes) will operate at different frequencies. The best results are obtained when the two channels are separated by a frequency difference of approximately 80 Mc/s; this is adequate to avoid mutual interference.

Single-user application. The Hydrodist system is not designed for multi-user application; each shore-based remote station is able to work with only one master instrument at a time.

Nevertheless, two and probably three vessels can operate (each with its own remotes) in the same general area provided that the operating frequency of each of the master/remote channels is selected so as to avoid mutual interference.

Future Developments for Hydrodist

The equipment referred to so far has reached an advanced stage of development. Acceptance tests and user trials are now being undertaken.

It will be noted that the masters previously described (Hydrodist Model MRB2) are portable and manually operated instruments suitable for either large vessels or small survey craft.

Current developments include the design of fixed or permanent installations for large survey vessels or dredgers. This equipment incorporates facilities for automatic readout of ranges, automatic ambiguity resolution, and automatic recordings of positional data.

The master installations, when available, will probably be 'purpose-built' equipments and will be designated Hydrodist Duplex Master Installation Model MRB/Duplex. They will normally constitute a single compact console unit, operated by one person, and will comprise two master measuring channels with an automatic digital readout of the two ranges. The signal output of the two measuring channels will also be capable of driving a two-range station plotting system such that an automatic plot of the ship's position is continuously given. In addition, the signal output will be capable of driving a graph recorder yielding a continuous and permanent record of the two ranges

phased with each other, and possibly phased also with an echo-sounding record. Another panel will provide for remote control of the bearings of the aerials.

Aerial rotation is effected from the console unit itself, either mechanically or through selsyn motors, as may be required.

The remote equipment used with the Duplex ship installation will be identical to standard Hydrodist Remote Model MRB2, except that the installation in permanent locations for unattended operation may present specific requirements as to siting, beam coverage, power supply, and weather protection.

5. AIRBORNE APPLICATIONS

General

The Tellurometer microwave system (in the form of Microdistancer equipment) has provided surveyors with a means of measuring lines accurately up to about 35 miles.

The use of aircraft obviously affords the opportunity to obtain line of sight over greatly increased ranges. If measurements can be made to an aircraft 100 miles distant from a ground point, it will be possible, by combining two such measurements, to measure up to 200 miles between ground points. By simultaneously making three such ground-to-air measurements the position of an unknown point in relation to two known points can be obtained.

Also, if the position of an aircraft in flight can be determined as above described, it may be possible to use such determinations as an aid in photographic survey. Similarly, it may be possible to employ aircraft positional determinations to assist in a topographic survey of an area or to find the position of an inaccessible point.

Aerodist is a Tellurometer system of line measurement under development for use in conjunction with aircraft. The work is being carried out under contract with, and to a purchase description of, the United States Army Engineer Research and Development Laboratories.

Aerodist, Theoretical Considerations

Theoretical considerations have been dealt with generally in section 3. The carrier fre-

quency of Aerodist land/air systems is in the 1200 to 1400 Mc/s band, and the size of the remote reflecting dish, the transmitter power, and the receiver sensitivity has been increased over the sizes of the standard Tellurometer system; these modifications extend the maximum working range. On the other hand, the directive gain of the master (airborne) antenna has been decreased, with a corresponding lowering of the maximum range.

The amount of gain obtainable from the reflector of the master antenna assembly is largely affected by considerations of the aerodynamic shape of the assembly and the necessity of having a wide beam to avoid continuous adjustment of direction.

The antenna system consists of the requisite number of crossed dipoles mounted on flat reflecting plates beneath the aircraft. The assemblies are mounted in pairs, back to back, each such pair being enclosed in a plastic bubble rotatable from inside the aircraft, so that if the aircraft changes direction by 180°, each dipole can be beamed onto its attendant remote. A duplex system needs only one double assembly; a triplex system needs two.

Pattern frequency. The master A pattern frequency is set at 1.49847 Mc/s. This frequency is selected to enable the instrument to read directly in meters, permitting a ready calculation of range. The value is devised on the assumption of a mean value of 1.0003 for the refractive index of the atmosphere. When high accuracy is required, continuous observations of atmospheric conditions are made from the aircraft.

The distance information is displayed on both a meter and a multichannel strip chart recorder. One channel of the recorder is reserved for each master/remote measuring channel. Since the measurement and display processes cannot distinguish between increasing and decreasing distances, an automatic ambiguity resolver is incorporated in the system. The ambiguity resolution for each master/remote channel is displayed on a recording channel alongside the relevant distance recording.

The remote (ground-based) units have high gain (and consequently narrow-beam) antennas. The remote operators must check the alignment of the antennas in azimuth and

just if there is any tendency for the aircraft to fly out of the beam. Pattern frequency switching is automatically carried out in such a way that the cyclic selection of pattern frequencies at the master is duplicated in all the slaves simultaneously.

Accuracy. The over-all accuracy of the Aerodist system is ± 1 meter ± 1 part in 100,000 the measured distance.

The 1-meter probable error is partly a function of scale resolution and partly a factor of the relationship between time and movement. It is probably the optimum that can reasonably be expected at speeds exceeding 100 miles per hour. At the same time this instrumental error becomes considerably greater significance at the shorter ranges likely to be measured (say 30 miles between ground point and aircraft). On the other hand (say distance of 75 miles between ground point and aircraft), the part affected by the basic instrumental error assumes lesser significance, so that under good measuring conditions the total error may quite possibly be less than 1 part in 100,000. In using the Aerodist system, this question of good measuring conditions does, in fact, present opportunities for accurate measurement not frequently found in normal measurements between land points.

Flight planning and ground stations. If the accuracy of the system is to be realized, restrictions must be observed as regards siting for flight planning, in particular aircraft height. These restrictions are interrelated. The siting of a ground station will often have to be determined by considerations other than those connected to obtaining optimum conditions for measurement. The aircraft height should not be kept to the minimum necessary to give a clear radio line of sight between the aircraft and the ground stations in the region where observations are to be made. Ground and station heights such that long path differences exist between the radio waves traveling directly between the ground and air stations and those reflected from the intervening ground should particularly be avoided. In general it will be found that these restrictions are of little significance at long range but that the use of the system at short range is subject to some restrictions.

It should be emphasized that this requirement of minimum suitable aircraft height is contrary in certain respects to general practice for UHF radio communications, in which the tendency is to arrange for the aircraft to fly at maximum heights. With Aerodist different considerations apply, and the tendency should be to fly at minimum necessary heights if the best accuracy and satisfactory signal strength and corresponding range are to be obtained.

Fortunately, this requirement is also one having distinct advantages from the aspect of reducing the measurements to geodetic data. The height of the aircraft is determined barometrically. Such determinations may be subject to errors (due, for instance, to diurnal or absolute deviation). The effect of such errors, so far as they concern reduction of the slant ranges to spheroidal (geodetic) distances, will be negligible when the aircraft is flying at low elevations but will increase if it is flying high. At the same time the aircraft should not fly so low as to render the barometric height unreliable because of ground turbulence or convection currents.

In the applications to aircraft hereafter suggested, the question of height has been considered from two aspects:

(a) The height difference between one end of a measured slope line and the other must be known. This is primarily required for the purpose of converting slope distance to horizontal (or spheroidal, etc.) distance. The accuracy of the height requirement for this purpose is dependent on the relationship between the height difference, on the one hand, and the length of the line, on the other. For measuring long lines with comparatively small height differences (or, for that matter, for positional determinations in the horizontal plane) the accuracy of the height determination is not critical, and barometric methods are nearly always considered to suffice.

(b) For position determination, however, the height component has a value in itself, and therefore consideration has been given to the accuracy that can be obtained by using line measurement to give an accurate height.

Ranging by Aircraft Line Crossing

An Aerodist duplex system comprises two

master channels in the aircraft operating in conjunction with two land-based remotes. It can be seen that this system will permit distances to be measured between two points on the ground as much as 200 miles apart with an accuracy of the order of 1 part in 100,000.

The requirement in the line crossing method is to obtain the minimum sum of the two ranges from the aircraft to the ground stations. The continuous range record permits the sum of the two distances to be plotted with respect to time and a graphical minimum to be deduced. By reason of reflection effects the sum of the two ranges will, at a particular point along the aircraft's flight, contain errors which are greater or less than the true sum. They will be small if careful attention is given to siting of ground stations and to aircraft height.

By graphing the sum of the two ranges the errors in individual readings are meaned out and a smooth curve is established. At the apex of the curve (which is the minimum sum) the errors due to reflection are partly, if not wholly, eliminated.

The minimum sum of the two ranges may be reduced to the corresponding spheroidal (geodetic) distance between the known and unknown point by means of established practices that take into account the properties of the reference spheroid and the respective heights of known and unknown points and of the aircraft. The height of the aircraft will have to be barometrically determined.

By arranging the flight plan so that several line crossings are made at different points it will further be possible to reduce considerably the error due to reflection effects inasmuch as the reflection characteristics at the actual time of one crossing are different from those of the other crossings. For each crossing a curve to show the minimum sum of two ranges is obtained, and these minimum sums are reduced to their spheroidal equivalents.

Position Fixing by Continuous Trilateration

An Aerodist triplex system comprises three master channels in the aircraft operating in conjunction with three land-based remotes.

The configuration of the ground stations is such that the aircraft can fly a course which keeps two ground stations to one side and one

to the other side. If different configurations are necessary, different antenna assemblies can be installed in the aircraft. By the recording technique which has been described above, a continuous record of three slant ranges is obtained. Each of these three ranges taken at a particular time during the aircraft's flight has errors due to ground reflection effects. During the flight of the aircraft, however, the path differences and reflection characteristics of the terrain change and the errors are contained in the small amounts by which the continuously recorded ranges are greater or less than their respective true ranges. By the selection of a number of points along the aircraft's flight and the reduction of the three slant ranges to spheroidal distances, a series of well conditioned trilateration figures can be built up and in effect the unknown point becomes the common apex of a number of self-supporting trilateral extensions derived from a series of bases of varying lengths.

Figure 1 gives a simple graphical representation of the method. The errors due to reflection rays which were inherent in the slant ranges will be contained in the relative spheroidal distances and hence also in a number of alternative positions for the unknown point. By this means the position of an unknown point as much as 200 miles from an established base can be determined in a single operation.

By taking a mean or adjusted value for the unknown point, and using a sufficient number of supporting figures, an accuracy of 1 part in 100,000 can be expected. If optimum accuracy is called for it is an advantage for the operation to be repeated on a flight plan which gives a different course well removed from the previous transit of the aircraft. Where special conditions exist, it may be advisable to repeat the transit at a different height.

It should be noted that by this method the height of the aircraft and of the unknown point will be barometric height, and the accuracy in that dimension of the unknown point will be a barometric function.

The position of two unknown points can be determined on successive flights. Also the position of the aircraft itself at any point on its flight can be determined from the recorded data.

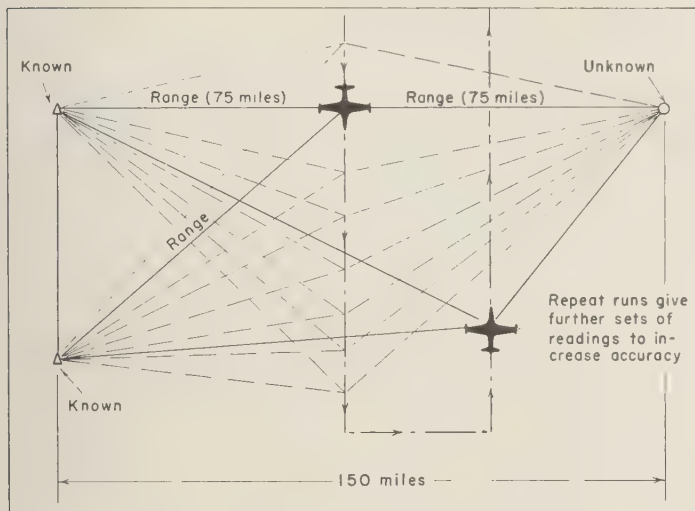


Fig. 1. Positioning by continuous trilateration.

Positioning an Aircraft

It is seldom that circumstances lend themselves to employing three line measurements as the sole means of the positioning of an aircraft. This is largely the result of considerations of practical application; for instance, if a large area is to be covered, in the location of the ground stations at known points primary consideration must be given to accuracy of positioning in the horizontal plane, and accuracy in the vertical plane has to be sacrificed.

Generally, for positioning of an aircraft over an area of wide coverage, barometric methods of height determination will normally be adopted and the position in the horizontal plane may then be supplied adequately by two line measurements only.

With two line measurements and barometric height, the aircraft's position can be determined by an Aerodist system to a range of approximately 100 miles from the base stations. Knowledge of the aircraft's position may be utilized in a number of ways.

Photographic Survey

The currently accepted procedure in photogrammetry relies on the determination of the spatial position of the camera relative to

known control points on the ground. The establishment of these control points entails considerable expenditure of time and labor; obviously a method that would allow the position in space of the aircraft to be determined at the instant of each exposure would obviate the necessity of establishing ground control except for the ground points to which the aircraft position is related. The possibilities of applying the principles of Aerodist to the problem have received preliminary consideration.

In photographic survey, the height factor must be determined with the accuracy that is called for to meet the requirements of scale and contour interval. When this is done ground heights can be obtained, from aircraft positional data, which are related to a height datum within the required accuracy, which in large-scale mapping with small contour intervals may be less than 1 foot. Barometric heights are unlikely to be sufficiently accurate for this.

An aircraft can be accurately positioned by Tellurometer line measurement from three ground points, two of which are remotely placed in order to obtain accuracy in the horizontal plane and the third of which is located more or less vertically below the aircraft. Normally the coverage area in a photographic survey operation is such that the horizontal components

of the aircraft's position at any instant can be determined with adequate precision by line measurements to two remote stations suitably located. A third line measurement to a suitably located remote station on the ground (or alternatively a barometric height) may provide the height component with sufficient accuracy for small-scale maps. The coverage area is likely to be too great to enable a single remote station to be located so that the vertical component of position can be determined with sufficient accuracy over the whole area for large- and possibly even medium-scale maps; it may, however, be possible for a pattern of points over the coverage area to be selected in such a way that remote units located at the points could, by operating singly on a time-sharing basis with the master unit in the aircraft, permit a sound geometrical figure for calculation of heights to be continuously obtained. The method can be employed so that the exact position of the aircraft is obtained at the instant of each exposure during the actual photo survey flights, or special control framework strips could be flown and the data thus obtained be applied by aerial triangulation to the actual photographs used for mapping.

For the height accuracies required for the larger-scale maps it would be necessary to modify slightly the master/remote channel which is used for measurement in the vertical plane, at the same time retaining the Aerodist system, as described previously, for horizontal positioning. The modification consists in changing the pattern frequency from about 1.5 Mc/s to a frequency more closely approximating that of the geodetic Tellurometer. By this means the required accuracy of measurement in the vertical direction can be obtained. Furthermore, antenna systems for vertical measurement would have to be suitably mounted.

Another possible application of Aerodist to photographic survey is for navigating an aircraft along definite flight lines, for which purpose facility for positional plotting in aircraft becomes a necessity. The practical problems have not yet been fully studied and evaluated.

Transferring Aircraft's Position to Ground Point

The advantages of the method for topographic mapping by aerial photography have

already been referred to. Whether the use of line measurements to an aircraft in flight may be applied so as to determine the position of points on the ground, otherwise than through the medium of aerial photography, remains to be discussed.

Provided that the difference in altitude between the unknown ground point and the aircraft (at the instant that the aircraft is directly over the point) is known, the ground point can be positioned. It is not the purpose of this paper to deal with methods of determining height difference, but available profile recorders and radio altimeters may be used within their prescribed limits of accuracy, and it may be well worthwhile to investigate the possibilities of other methods. Such a procedure would necessitate accepting the recorded height of the aircraft as the height datum. As stated elsewhere in this paper, it is probable that, for large area coverages, practical considerations would normally result in reliance on barometric aircraft height to provide the height datum, with the consequent inaccuracies inherent in the method.

A simple optical sighting device may possibly be used to determine the exact time when the aircraft is in fact over the ground point, and a 'event marker' on the recorder may be employed so that the position is identified.

Assuming that information relating to aircraft position can be transferred to a ground point with appropriate accuracy, it is easy to see that, by this means, an extension or breakdown of the existing network of survey beacon over an area of large coverage could be undertaken in a very much shorter time than by present ground-survey methods. In addition the procedures proposed may be particularly applicable to the positioning of an inaccessible point.

6. CONCLUSION

In presenting this paper on new applications to geodesy and hydrography it must be appreciated that much work still remains to be done, particularly on the airborne application. It was my intention to try to outline some of the possibilities and the problems involved in the new applications now envisaged.

Doubtless great advantage will accrue from discussion and deliberation, and methods of im-

ed techniques of application will very likely
orthcoming.

acknowledgments. The work that provides
basis for this paper was carried out at the Re-
sh and Development Laboratory of the In-
nent Manufacturing Corporation of South
ca, Ltd., in Cape Town.

knowledgment is made to the Director and
of the National Institute of Telecommunica-
Research (S. A. Council for Scientific and

Industrial Research) for their invaluable assis-
tance.

REFERENCES

- Wadley, T. L., The Tellurometer system of dis-
tance measurement, *Empire Survey Rev.*, 24,
nos. 105 and 106.
Wadley, T. L., Electronic principles of the Tel-
lurometer, *Trans. S. African Inst. Elec. Engrs.*,
47, 143, 1958.

Report on Electronic Distance Measurements in Australia

G. R. L. RIMINGTON

*Division of National Mapping
Department of National Development
Melbourne, Australia*

Australia is a large, comparatively undeveloped country, with a small population, as the following statistics show:

Area	2,974,581 square miles
Coastline	12,210 miles
Population	10,000,000

Its present state of development finds only a small amount of geodetic survey completed, and such a small population cannot provide much manpower to cope with the task of establishing a geodetic survey. Consequently, all methods of increasing the effectiveness of available staff must be explored.

The first investigations into the use of electronic distance-measuring equipment were made by the Radiophysics Division of the Commonwealth Scientific and Industrial Research Organization during 1947-1949 [Warner, 1950].

At the request of the National Mapping Council, the Division examined the possible application of Shoran to geodetic surveying. The sides and diagonals of a quadrilateral were measured; they varied in length from 158 to 311 miles. The resultant measurements were compared with distances computed from a first-order ground triangulation. The results of more than 100 measurements spread over 6 lines gave an over-all accuracy of about 7 parts in 10^6 . The scatter of the individual measurements on any line was about one-third of this. Systematic equipment errors proved the main limiting factor in obtaining higher accuracy. The greatest source of error in the radar equipment was associated with signal strength.

On completion of its experiments the Radiophysics Division advised that by suitable modification to the equipment, in particular the receivers, this error in signal intensity could be reduced to within ± 0.002 mile. If this were done it would be likely that the over-all ac-

curacy of the technique would improve to about 2 parts in 10^6 . An improvement on this figure would be impossible without extensive improvements in the equipment and a thorough investigation of problems of atmospheric refraction.

At about this time Dr. Bergstrand had demonstrated his first models of the Geodimeter, and it was decided to acquire and test a production model of the instrument before finally deciding on a type of electronic distance-measuring equipment for use in Australia.

A unit of the equipment was obtained in 1953 and thoroughly tested [Waller, 1954], with results that have had great significance. The results of these tests, which agreed consistently with those carried out in other countries of the world, established that electronic measurement of the lengths of normal geodetic lines was not only practicable but also had a phenomenal degree of accuracy. It was considered a remarkable advance over Shoran airborne electronic measuring equipment, which relies on extreme length of line (up to 400 miles) to achieve its fractional accuracy.

It was established that the Geodimeter Type NASN.1 could measure distances with a limiting error of ± 0.08 ft. Here then was a system that could eliminate the time-consuming base line and its associated network, substituting the direct measurement of a single side of triangulation for all such base lines.

Additional Geodimeters Type NASN.2 were obtained and placed in use in Australia, on the measurement of base lines for first-order triangulation. This equipment for maintenance of scale in classic triangulation proved so successful that its use for making measurements on first-order traverse was inevitable.

Some small schemes of traverse were actually completed, with satisfactory results, leading to

the conclusion that, where the topography allowed the use of the Geodimeter, traversing could easily meet first-order accuracy, with considerable saving in manpower and time.

Plans were being prepared for extensive application of Geodimeter traversing when the Tellurometer made its appearance. Preliminary reports on the Tellurometer system of electronic measurement indicated that it was intended to carry out much the same task as the Geodimeter, the basic difference being that the Geodimeter used light waves as the measuring medium and the Tellurometer used radio waves. The specifications of the Tellurometer were such that it offered substantial advantages over the Geodimeter when used on traversing.

Australia was not alone in appreciating the possibilities of the equipment; we in this country were one of the many survey authorities which obtained early production models for investigation.

Our aims in this testing were perhaps slightly different from those of other organizations, in that we had been conditioned by exacting tests of the Geodimeter into accepting the conclusion that first-order traverse carried out by electronic equipment was the most economical solution to our pressing needs in geodetic survey. The testing of the Tellurometer was therefore aimed at determining the relative accuracy and efficiency of the Geodimeter and Tellurometer when both systems were operating at maximum efficiency.

The results of the tests of the Tellurometer were published in Australia [*Rimington*, 1957], and limited copies of the report were distributed to interested overseas organizations; the report concluded that the results of Tellurometer measurements were very little inferior to those obtained with the Geodimeter.

It is conceded that in some extreme circumstances the Tellurometer will fail to measure accurately between two fixed points, and that the use of short eccentric observing points may be necessary to avoid the effect of 'ground wings.' Experience has indicated that wide wings which may cause inaccuracy can be minimized by a small change in the instrument standpoint. In support of this belief, it is on record that in some 9000 miles of accurate meas-

urements it has always been possible to measure satisfactorily all previously selected lines.

EFFICIENCY

The physical characteristics of the Geodimeter and Tellurometer are such that it is very easy to decide which of the two instruments is preferable for extensive field work. Weight, cost, and susceptibility to weather conditions all weigh so heavily against the Geodimeter that all authorities engaged in geodetic survey in Australia are using Tellurometers in the field. The few Geodimeters are held in reserve for occasions when extreme accuracy of the order of ± 2 parts in 10^6 is required, or when the line is so short that the limiting accuracy of the Tellurometer (± 0.20 ft) cannot be tolerated. Such occasions are very rare.

FUTURE PLANNING

On present planning the initial stage of the Geodetic Survey of Australia will take the form of extensive loops of Tellurometer traverse encircling and subdividing the continent.

It is realized that there may still be a considerable difference of opinion as to the accuracy and dependability of the Tellurometer, and it is the purpose of this report to set out some of the results of the testing and work carried out in Australia during the past two years. These results have bred confidence in electronic measurements to such an extent that a major change in methods of geodetic survey is accepted in Australia.

USE OF TELLUROMETER FOR GEODETIC MEASUREMENTS

General considerations. Measurements made with this instrument are subject to: (a) index error; (b) systematic error in the accepted velocity for electromagnetic waves in vacuo; (c) variation of frequency of the crystals in the equipment; (d) inaccuracy in the accepted meteorological corrections; (e) errors in measurement due to reflected waves (swing).

Index error. This type of error undoubtedly exists in every measurement made with the equipment. The designer informs us that it will always be of the same sign (distance measured too long) and can vary between 0.00 and 0.25 ft.

TABLE 1. Tellurometer calibrations, Balcombe, Victoria
Measurements are in feet.

Tellurometer and Index Correction Used		MA.11, -0.23		MA.40, -0.19		MA.74, -0.25		MA.318, -0.0	
Line	Taped Feet	Meas- ured	Residual	Meas- ured	Residual	Meas- ured	Residual	Meas- ured	Residual
W. base-E. base	5899.45	5899.50	+0.05	5899.50	+0.05	5899.52	+0.07	5899.56	+0.0
Peg 18-E. base	2951.37*			2951.31	-0.06	2951.33	-0.04	2951.36	-0.0
Peg 18-W. base	2948.08*			2948.18	+0.10	2948.01	-0.07	2948.10	+0.0
Sum (E.-W.)	5899.45			5899.49	+0.04	5899.34	-0.11	5899.46	+0.0
Peg 19-E. base	3115.44	3115.48	+0.04	Note: Peg 19 was destroyed after first calibration					
Peg 19-W. base	2784.01	2784.05	+0.04						
Sum (E.-W.)	5899.45	5899.53	+0.08						
Peg 20-E. base	3279.34	3279.31	-0.03	3279.30	-0.04	3279.21	-0.13	3279.12	-0.2
Peg 20-W. base	2620.11	2620.10	-0.01	2620.06	-0.05	2620.28	+0.17	2620.25	+0.1
Sum (E.-W.)	5899.45	5899.41	-0.04	5899.36	-0.09	5899.49	+0.04	5899.37	-0.0
Total of E.-W. residuals			+0.09		0.00		-0.00		+0.0
Mean of E.-W. residuals			+0.03		0.00		-0.00		+0.0

* Provisional, may be ± 0.02 ft in error.

Being always of the one sign, the error will be cumulative in traverse work, and some effort should be made to ascertain and apply a correction. It is difficult to determine a correction of this order of magnitude when: (a) the smallest scale division is 0.50 ft; (b) the presence of swing may cause errors of the order of ± 0.20 ft; (c) completely reliable reference lengths are required; (d) overloading of the equipment must be avoided.

The Division of National Mapping has been using a test site at Balcomb, Victoria, which has fairly satisfactory characteristics. It is a former training base line of the Royal Australian Army Survey School. The line is 5899.45 ft long, over slightly undulating country, which at certain times of the year has a covering of long grass. When long grass covers the site, ground swings are at a minimum; when the grass is dry and short, only moderate swings are developed. Intermediate pegs are available along the length of the line, and it is possible to measure the line in whole and in parts.

It is of particular importance that all measurements with the original model of the Tellurometer were carried out with the crystals at 'turn-over' temperature (see under 'Frequency,' below). The crystals were calibrated in such a manner that the turn-over temperature was accurately determined.

Table 1 shows the results of the tests for index error for equipment numbers MA11, MA40, MA74, and MA318. It will be observed that the results of the MA11 and MA40 tests are noticeably more consistent than those of MA74 and MA318. The former two tests were made with long grass covering the line, and the latter two with short grass cover.

These results clearly show both the existence of an index error and its probable amount. They also give some idea of the consistency of the instrument at short range under favorable conditions. Tests such as these confirm the conclusion arrived at after our early tests, that the manufacturers' specification of accuracy is substantially correct at distances of this order.

Velocity of electromagnetic waves. The present accepted velocity of 299792.5 km/sec is the result of experimental work carried out by different methods, and has been endorsed by the I. U. G. G. The use of this value should not introduce a systematic error in excess of 1 part in 600,000. An error of this magnitude is insignificant in normal geodetic work. It is notable that the velocity deduced in various countries of the world from the use of the Tellurometer over accurately measured lines shows an amazing agreement with this figure. It is very strong contributory evidence that the equipment is capable of a high order of accuracy.

Frequency. The initial models of the Tellurometer relied on crystal monitoring to determine the actual frequency used during a measurement. To obtain a high order of accuracy in frequency it was necessary to operate the equipment at the 'turn-over' temperature of the crystal, at which temperature errors in the monitoring circuit were at a minimum. By means of such a technique, combined with frequent reference to standard frequency, it has been possible to preserve frequency stability to within 1 part in 10^6 .

A modification is available in which the crystals are maintained in a controlled oven. Frequency is maintained within 1 part in 10^6 without special routines. All equipment being used in Australia on geodetic survey is being fitted with this modification.

Meteorological corrections. This is perhaps the most controversial issue in any discussion of the Tellurometer. The formulas on which the corrections are based appear to be acceptable for the order of accuracy envisaged in geodetic survey, and it is not anticipated that they will be seriously challenged for validity.

Debate generally centers around the accuracy of the conventional meteorological measurements, and the propriety of using readings made at both ends of a line as representative of conditions over the whole line. Bearing in mind the vagaries of atmospheric conditions it is understandable that this aspect of the measurement should be viewed with suspicion. Perhaps this skepticism about the meteorological corrections is felt by all who come into contact with electronic distance measurements. In the initial tests of the Geodimeter and Shoran much the same uncertainty prevailed, but as further work progressed it was replaced by confidence.

The Tellurometer is more sensitive to vapor content of the atmosphere than the Geodimeter, but, as the results of many measurements show, agreement under widely varying conditions, confidence in these corrections will be established.

In work of geodetic quality it is proper that due regard should be paid to the chance that freak atmospheric conditions may lower the quality of the measurement; to guard against this possibility, geodetic lines are measured twice, once on each of two different days. This procedure is an insurance both against freak

meteorological conditions and against the possibility of gross errors.

A tabulation of the dual readings reveals interesting information, particularly when the meteorological corrections are compared. A small sample appears as Table 2, listing the lines of a traverse running from Alice Springs (vicinity) to Halls Creek in Central Australia. The temperatures experienced on this particular survey varied from 64°F to 99°F , and the results indicate that, despite violent changes of temperature, the agreement between the two measurements of the lines is satisfactory. The line Gardiner-Finniss (last measurement) failed in this respect and will be remeasured. This tabulation is typical of the agreement obtained over lines measured in Australia; it has been a rare occurrence for two measurements to fail to agree within 1 part in 10^5 . In the event of such disagreement the line is remeasured.

Ground swings. Much has been written about 'ground swings,' the term applied to the measured effect of reflected waves on the direct wave. That such ground swings exist is obvious, and their presence on very short lines is pronounced.

Almost without exception, the new user of Tellurometer equipment will first test it over very short lines, with the inevitable result that quite large ground swings are encountered, which have a catastrophic effect on the proportional accuracy obtained.

The situation changes completely when the equipment is placed in field use, measuring lines of main-scheme geodetic control. With the ruling length of line in the vicinity of 20 miles and over, it becomes quickly apparent that ground swings exceeding $6\text{ }\mu\text{sec}$ (or 3 ft) are few and far between. Referring to Table 2 again, it is seen that the normal amount of ground swing is of the order of 2 or 3 μsec . In a line of length 20 miles the whole extent of the ground swing represents approximately 1 part in 60,000; when the ground swings have been plotted and analyzed in the usual fashion, their effect on the measurement is not considered significant.

Recommendations. Earlier this year, the National Mapping Council of Australia, after reviewing the development work on the Tellurometer and the uniformity of results obtained by various agencies, endorsed the following advice

TABLE 2. Tellurometer Traverse, Alice Springs to Halls Creek
Backward and forward measurements made by the Division of National Mapping during the 1958 field season are compared.

Slope Distance Eccc to Eccc	Forward, ft	Meteoro- logical Correction	Range of Swings, ft	Backward, ft	Meteoro- logical Correction	Range of Swings, ft	Difference,		Measurement 1 : Fractional	Difference in Meteor. Corr.	
							Forward	Backward, ft		Forward	Backward, ft
Gardiner-Treachery	127,420.55	32.02	3.37	127,421.04	35.26	2.87	-0.49		262,000	-3.24	
Treachery-Matthews	57,846.50	15.28	3.4	57,846.61	14.66	3.5	-0.11		527,000	+0.62	
Campbell-Matthews	121,714.77	31.39	1.4	121,714.91	32.07	1.1	-0.14		860,000	-0.68	
Campbell-Patricia	162,618.46	42.08	1.4	162,617.88	43.01	0.9	+0.58		279,000	-0.93	
Theo-Patricia	78,201.35	21.31	4.5	78,200.82	21.05	3.7	+0.53		148,000	+0.26	
Theo-Bennett	166,821.84	42.75	1.2	166,822.76	43.25	0.6	-0.92		181,000	-0.50	
Bennett-Solitaire	137,222.70	36.19	1.0	137,223.55	37.12	0.9	-0.85		161,000	-0.93	
Solitaire-Davidson	113,178.28	29.39	2.0	113,178.06	30.56	2.3	+0.22		513,000	+1.17	
Davidson-Approach	151,969.91	40.12	0.9	151,910.38	40.55	1.2	-0.47		321,000	-0.43	
Approach-Granites	48,406.03	12.85	1.2	48,405.90	13.11	1.2	+0.13		371,000	-0.26	
Granites-Pilotus	117,050.50	31.54	2.5	117,050.75	31.55	3.2	-0.16		728,000	-0.01	
Pilotus-Frankenia	112,973.30	30.10	1.5	112,973.43	32.19	1.4	-0.13		677,000	-2.09	
Frankenia-Tanami	103,627.96	29.32	1.5	103,627.90	27.75	1.4	+0.21		490,000	+1.57	
Tanami-Frederick	184,945.47	49.76	1.7	184,946.78	49.93	1.6	-1.31		141,000	-0.17	
Frederick-Junction	179,210.58	46.92	1.0	179,211.31	47.01	0.6	-0.73		245,000	-0.09	
Junction-Westwall	180,962.08	48.77	2.1	180,962.33	48.90	1.8	-0.25		722,000	-0.13	
Westwall-Browns	45,274.58	12.33	0.9	45,274.41	12.15	0.7	+0.17		267,000	-0.08	
Browns-Windoo	127,126.74	36.71	1.3	127,126.61	34.94	1.0	+0.13		975,000	+1.77	
Oaks Windoo	128,598.21	41.69	1.5	128,590.47	33.46	0.5	-0.26		495,000	+8.23	
Gardiner-Finniss	146,352.07	35.81	3.1	146,354.44	38.78	2.7	-2.37		62,000*	+2.97	

* This line will be remeasured.

tendered by its Technical Subcommittee, and recommended that it be taken into account in future Geodimeter and Tellurometer operations:

The Subcommittee believes that, after consideration of the work carried out during the past two years, an average accuracy can be obtained with electronic distance-measuring equipment as follows:

- a) Geodimeter, 2 parts in 10^6
- b) Tellurometer, between 5 and 10 parts in 10^6

Recommended distances for use of various instruments for geodetic purposes:

- Geodimeters, from 1 to 20 miles
- Tellurometers, from 10 to 50 miles

Frequency. Recommended that error in frequency of the crystal should not exceed 1 part in 10^6 .

The Subcommittee is keenly aware of the danger of possible changes of frequency in the field and is of the opinion that simple and economical equipment can be designed to carry out field checks of the frequency of the 'A' crystal. It recommends that action be taken to have such equipment designed and tested.

Ground swing with Tellurometer. The Subcommittee has considered the effects of ground swing on measurements, and believes that the treatment suggested by Mr. Wadley in the handbook of the equipment will provide a satisfactory interpretation of such swings.

Meteorological corrections. The Subcommittee is of the opinion that the formulas for meteorological corrections, as laid down in the operating handbook, will give results which are acceptable for the order of accuracy envisaged in paragraph 1 of this recommendation.

Recommended practices in operation of the Tellurometer. Amplitude of ground swings. The Subcommittee is of the opinion that the recommended accuracy can be obtained from measurements in which the regular ground swing does not exceed 6 μ sec in total transit time.

Where an initial measurement discloses ground swing of greater than 4 μ sec it is desirable to

change the standpoint of the instrument in an endeavor to reduce this ground swing.

A ground swing of 10 μ sec or more is considered to be a breakdown in measuring technique for geodetic purposes, and in such cases relocation or division of the line should be considered.

In continuation of conventional survey practice it is recommended that all first-order control measurements be carried out twice, each measurement covering the whole range of cavities.

PROGRESS TO DATE

During the past 1½ years approximately 40 Tellurometer systems have been, or are in the course of being, placed in service in Australia by Commonwealth and State Survey authorities.

In addition to geodetic survey, a large amount of work is being carried out as control for topographic mapping.

With the geodetic survey in its skeleton form, many of the topographic measurements are made over lines that may at some future time be incorporated in the main geodetic scheme.

In view of the comparative ease with which first-order measurements can be made with the Tellurometer, it is standard practice to read all long lines to the recommended standard of accuracy of 1 part in 10^5 .

The approximate total length of lines measured to this order of accuracy to date is 9000 miles, and it is apparent that within 5 years the Geodetic Survey of Australia will be well on the way to completion.

REFERENCES

- Rimington, G. R. L., Report on Tellurometer tests, *Australian Surveyor*, 16, 494, 1957.
- Waller, C. K., Field use of the Geodimeter, *Cartography*, 1 (3), 1954.
- Warner, J., The application of radar to geodetic surveying, *Australian J. Appl. Sci.*, 1 (2), 1950.

Tellurometer Measurements in the Base Extension Network Munich

R. SIGL

*German Geodetic Commission
Munich, Germany*

The development of instruments and of procedures for distance measurements by means of high-frequency-modulated light or by means of electromagnetic waves has made significant progress within the past few years. Since Germany has had but little share in this work and in testing the new geodetic methods of measuring, we wanted to present for the first time practical experience with the Tellurometer measurements reported here, particularly about: (1) fixation of the network by means of pure distance measurement; (2) determination of the possible accuracy for trilateration lines and for the point positions; (3) investigation of influence of meteorological conditions and of ground formation on the measurement of the transit time of electromagnetic waves.

The measurements were performed in 14 days in the autumn of 1958. All the lines of the extension network (Fig. 1) and a second line of the primary triangulation were measured; the lengths were 8.2 to 56.3 km. To determine changes, if any, of the instrument, one line (base line north-base line south) was measured at the beginning, in the middle, and at the end of the observations. The complete measurement of the first extension triangle showed no differences between the measurements in opposite directions; hence further return observations were not made, in view of the homogeneous ground conditions in the whole network. On all stations four complete measurements were carried through. It was found that the best observation conditions, and consequently the time requirements, depend primarily on the weather and on the time of day. Overcast, mist, and rain are all favorable; with bright sunlight satisfactory measurements were possible only in the morning and in the late afternoon.

The oblique distance D' was derived from

the measured transit times and the meteorological data in the usual way, according to

$$D' = \frac{1}{2} \theta \cdot c \cdot (1/n) \quad (1)$$

with

$$(n - 1) \cdot 10^6 = 103.46 / (273 + T) (B + K) \quad (2)$$

$$K = 4744 / (273 + T) \cdot e \quad (3)$$

where

B = air pressure, mm Hg.

e = vapor pressure of air, mm Hg.

T = air temperature, °C.

The vapor pressure e was derived from

$$e = E - 0.48 \cdot (T - t) / (610 - t) \cdot B \quad (4)$$

where E is the saturated water vapor for the wet-bulb temperature t . For the light velocity in vacuo the value 299,792.5 km/sec recommended by the International Association of Geodesy was used. As far as accuracy of the oblique distances is concerned, it made no difference whether the repeated measurements of one line were performed on the same day or on different days, in different directions, or with highly different atmospheric conditions; the results differed only by a few centimeters. A graph of the mean errors computed from the repeated measurements shows no apparent dependence of the mean errors on the distances; the maximum mean errors occur with the measurements performed on sunny days.

One line, Munich to base line south, caused great difficulties. The precise readings oscillated highly ($\Delta_{\max} = 5 \cdot 10^{-9}$ sec \cong 0.35 meter), and it was found that the available range of the carrier frequency was too small for the measurement (the carrier frequency of 3000 Mc/s

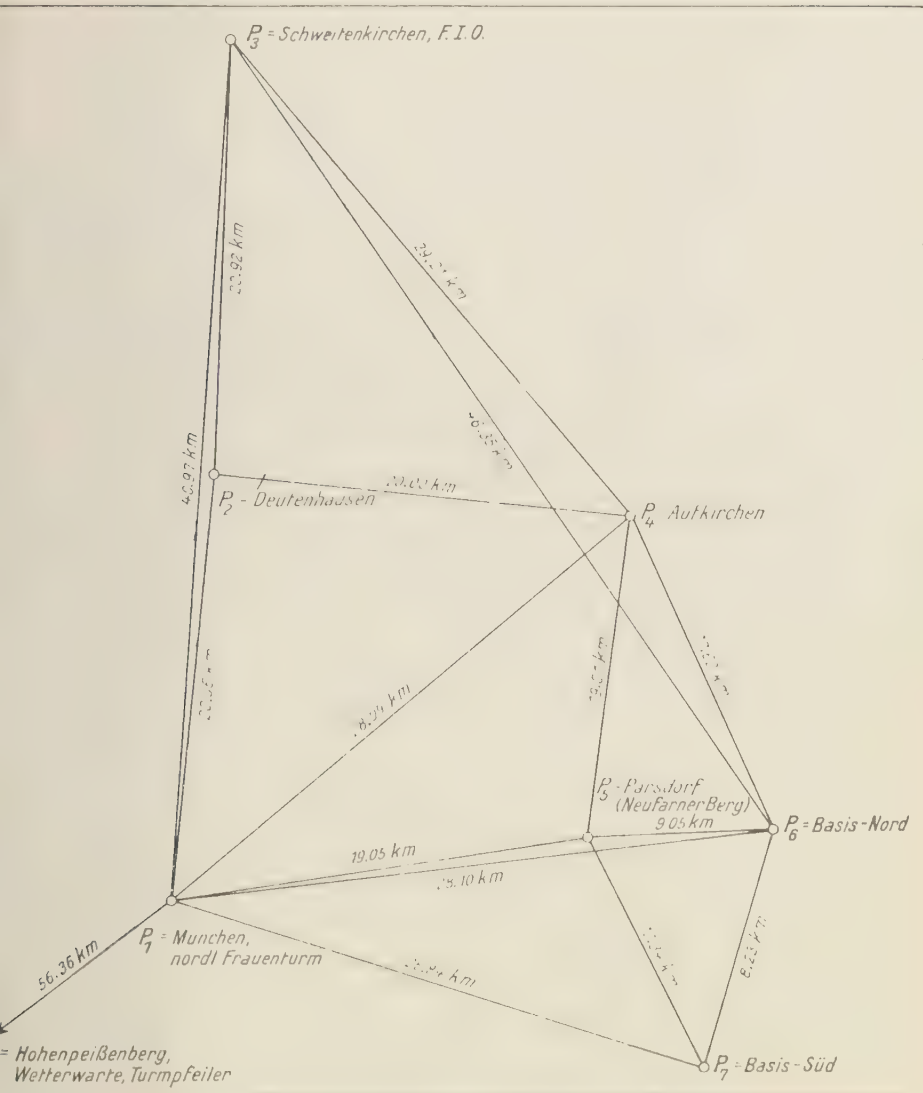


Fig. 1. Base extension network Munich. Scale 1 250,000.

ried in 12 steps over a range of ± 10 Mc/s).
 explanation for this has been found; an
 adjacent line with similar ground contour could
 be measured without difficulties under the same
 rather conditions.

The following expression is sufficiently accu-
 e for the reduction of the measured oblique

distances D' to the computation reference sur-
 face:

$$S = \sqrt{\frac{(D' + \Delta H)(D' - \Delta H)}{(1 + H_1/R)(1 + H_2/R)}} + \frac{(D')^3}{24R^2} \quad (5)$$

where

ΔH = difference of elevation.

TABLE 4

Side	Tellurometer Measurement (1)	Bavarian Primary Triangulation (2)	Base Extension Network (3)	Differences	
				(2)-(1)	(3)-(1)
...-Schweitenkirchen	40974.775	40974.554	40974.764	-0.221	-0.011
...-Hohenpeissenberg	56364.225	56363.93	...	-0.295	...

TABLE 5

Side	Plane Side Lengths, m		Corrections for Adjustment, m			
	Measured	Equalized according to Equation 11	I	II	III	IV
1.2	20 056.975	056.855	-0.054	-0.049
1.3	40 975.237	975.016	-0.012	-0.003	+0.042	+0.036
1.4	28 090.262	090.033	+0.012	+0.024	+0.012	+0.024
1.5	19 053.527	053.411	-0.038	-0.038	-0.038	-0.038
1.6	28 100.454	100.295	+0.029	+0.015	+0.029	+0.015
1.7	26 838.807	838.654	-0.008	-0.001	-0.008	-0.002
2.3	20 918.424	918.299	-0.005	-0.043
2.4	20 003.804	003.584	+0.001	+0.005
3.4	29 206.183	206.018	-0.052	-0.059	-0.058	-0.066
3.6	46 348.713	348.466	+0.070	+0.066	+0.074	+0.067
4.5	17 009.573	009.467	-0.002	+0.006	-0.002	+0.007
4.6	17 623.069	622.960	-0.070	-0.068	-0.090	-0.068
5.6	9 047.662	047.594	-0.044	-0.034	-0.044	-0.035
5.7	10 337.583	337.509	+0.010	-0.002	+0.010	-0.002
6.7	8 231.927	231.863	-0.009	+0.003	-0.008	+0.004

ent there were used the directly measured
e lengths and those equalized to the primary
angulation in accordance with

$$v_{mm} = -24 - 4.81 \times S_{km} \quad (11)$$

For the adjustment of the trilateration network
was at first necessary to select a proper com-
utation reference surface. The network was
signed in a plane mapping (Gauss-Krüger
jection) because it became possible to compute
efficiently precise coordinates for all the points
the network and hence rigorous side reductions
s. This implies essential simplifications
ainst the otherwise necessary computation on
e spheroid or on a Gauss sphere. Thus the
asured side lengths mean the planed values
f the spheroidal side lengths S .

$$s_{i,k} = S_{i,k} + \Delta s_{i,k} \quad (12)$$

The following considerations were decisive for
the selection of the adjustment method: Cor-
responding to the $r = s - (2p - 3) = 4$ surplus
sides there result four normal equations with
adjustment by conditioned observations but
eleven normal equations with adjustment by
variation of coordinates. The time factor was
not decisive for the selection of the method,
because a program-controlled relay computing
machine Z11 was available for the solution of
the normal equations. Thus the adjustment by
variation of coordinates was preferred for the
following reasons: (a) the procedure is easier
to survey, and the error equations are simpler
to establish than the condition equations; (b) the
result is given directly in terms of the coordinates
of the new points, whereas otherwise they can
only be computed from the adjusted side lengths.

The error equation in terms of plane coordi-

nates has the form

$$\begin{aligned}
 v_{i,k} = & \Delta y_i \sin \bar{t}_{k,i} \\
 & + \Delta x_i \cos \bar{t}_{k,i} + \Delta y_k \sin \bar{t}_{i,k} \\
 & + \Delta x_k \cos \bar{t}_{i,k} + \bar{s}_{i,k} - s_{i,k} \quad (13)
 \end{aligned}$$

where \bar{t} and \bar{s} are the values for direction angle and side length derived from the approximate (adopted) coordinates. All the measured sides were taken with the same weights ($p_{i,k} = 1$).

The adjustment was performed in four different ways:

- | | | |
|--|---|---|
| I. Measured side lengths | } | Form of the base extension network without station Deutenhausen |
| II. Side lengths equalized to the primary triangulation according to equation 11 | | |
| III. Measured side lengths | } | Complete network |
| IV. Equalized side lengths | | |

Only the side lengths and the corrections ensuing from the adjustments are given as results, because they indicate the intrinsic

accuracy of the Tellurometer measurement. See Table 5.

On the average the corrections for the measured sides amount to ± 3.5 cm. Thus the geometrical network conditions are well fulfilled; gross systematic errors do not occur. Furthermore, the correctness of the distribution of weights is proved; large and small corrections occur with the same frequency for the long and short sides. Thus accuracy is not a function of length. Such further factors as ground or unfavorable atmospheric conditions which are reflected in the deviations of the individual measurements could evidently not be used for assignment of weights. The mean error of a measured side after adjustment is ± 0.07 meter. Comparison with the mean errors derived from repeated observations suggests an additional error affecting the measurements, which, however, cannot be explained. The average position accuracy of the new points fixed by pure distance measurements amounts to ± 0.1 meter. Evaluation of the external accuracy and eventual scale errors will be reserved for later comparison with the results of base-line measurement and triangulation.

Results of Tellurometer Measurements in 1958

KARL GERKE

*Institut für Angewandte Geodäsie
Frankfurt, Germany*

REMARKS

The Institut für Angewandte Geodäsie, Frankfurt/Main, Section II of Deutsches Geodätisches Forschungsinstitut, purchased Tellurometer equipment in March 1958. Numerous measurements of test lines and standard lines have been carried out in the areas of Frankfurt/Main, Hanover, Münster, and Munich in close cooperation with the competent land surveying agencies during the period of field work in 1958. The measurements and results will be reviewed in a volume of the publication series of the German Geodetic Commission. A summary of the review is presented here.

MEASUREMENTS

The measurements have been executed according to the normal measuring method with the complete Tellurometer equipment consisting of one master station, one remote station, two barometers, and two hygrometers.

1. Area of Frankfurt/Main. The first measurements in the area of Frankfurt/Main served for training the observers. A total of 9 single lines with lengths varying from 1 to 30 km, selected according to various profiles of the terrain and to different types of coverage of vegetation, were measured several times under different meteorological conditions. One line of 1 km length has been determined in such a way that 3.5 km of it passes about 2 to 3 meters above the level of the river Main. This line offers good measuring conditions; the results did not indicate unusual deviations.

2. Area of Hanover. In the area of Hanover standard lines have been measured in the field Empelde, an excellent trigonometric network of the Niedersächsisches Landesvermessungsamt. There are 30 lines with an average length of 1.8 km in a traverse net, with longer lines and 19 lines in a closed net which is part

of the triangulation network of the city of Hanover. The city network also includes the geodetic base line of Meppen with 2 lines of the expansion network.

Several lines have been measured placing the instrument from 1.5 to 13.5 meters above the ground. With increasing elevation above the ground shorter transit times were measured. The amplitudes of the fine reading curve decreased, and the course of the curve became steadier. The shorter values here were in better agreement with those computed from the good trigonometric basic data. Several lines could not be measured with normal tripods.

Many of the traverse lines which had been measured only once have been checked with a 100-meter tape.

The trilateration network of Hanover between the stations of the triangulation network has been measured several times. The scale of the triangulation network available for the comparison is determined with unusual precision. The mean errors of the adjustment of the trilateration network will be mentioned under section 5.

The base line of Meppen with its slightly concave ground configuration and its sandy soil offers nearly perfect measuring conditions. The measuring ray of the primary triangulation line Windberg-Hesepe derived from this base line passes very closely above a forest.

The measurements of the distances between the primary stations of the triangulation network Brillit, Sievern, and Wingst have been carried out on the trigonometric high signals. The angles computed from the Tellurometer results are in very good accord with the primary angular measurements of 1955-1956 (see table on next page).

3. Area of Münster. In the area of Münster 58 traverse lines, with an average length amounting to 1.5 km, and 2 lines of the primary triangulation network have been measured.

Angle	Tellurometer	1955-1956	Difference
Brillit-Sievern-Wingst	63°46'17.318"	63°46'17.738"	-0.420"
Sievern-Wingst-Brillit	67°43'11.410"	67°43'12.025"	-0.615"
Wingst-Brillit-Sievern	48°30'33.765"	48°30'34.040"	-0.275"
Σ	180°00'02.493"	180°00'03.803"	-1.310"

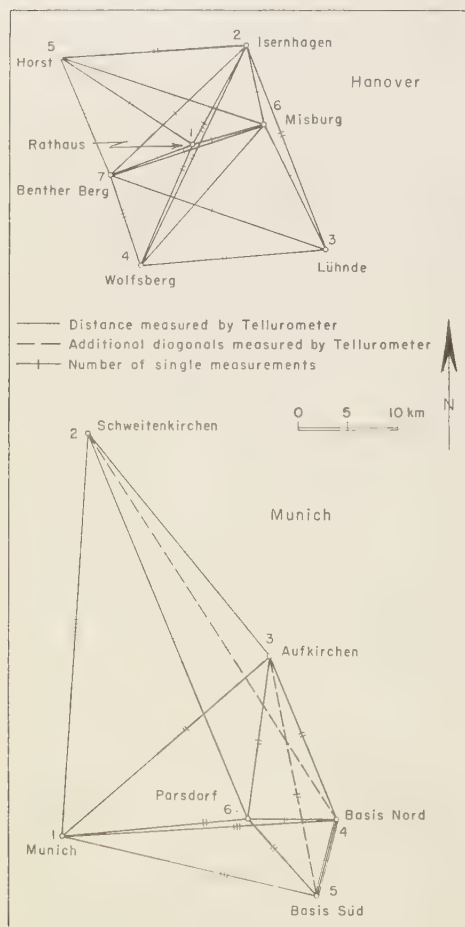


Fig. 1. Tellurometer measurements in 1958, Hanover and Munich.

As the traverse lines were measured only once, the results have not been considered in the computation of errors. They indicate that measuring of traverse lines in this flat area with normal timber is rather difficult. Small strips

of wood and windbreaks made measuring impossible, and even at longer sights the necessary for traversing they complicated the measurements. In general deciduous trees cause more trouble than conifers.

The measurements of both of the primary triangulations of 50 and 63 km length show that this is the extreme range of the instrument; the reading circle appeared fuzzy, and the windbreaks were hardly visible.

4. *Area of Munich.* All Tellurometer measurements in the extension network of the base line Ebersberger Forst near Munich, measured with Invar tapes in 1958, were executed with high signals. Most of the lines were measured twice with exchanged instruments, but the base line was measured 9 times and the derived primary triangulation line Munich-Schweitenkirchen 5 times. The repeated measurements of the lines Munich-Schweitenkirchen and from Munich to the southern base-line terminal, as well as the baseline between the terminals north and south, showed no essential deviations compared with preceding measurements after the installation of the arctic equipment. A comparison with lines trigonometrically determined will be possible only after the adjustment has been completed.

5. *Adjustment of the trilateration network.* The trilateration networks of Hanover and Munich (Fig. 1) have been adjusted by means of observation equations as well as by condition equations, introducing different weights. For both networks the following weights have been selected:

Hanover

$$p_1 = 1$$

$$p_2 = n$$

$$p_3 = 1/c$$

n = number of measurements.

c = largest amplitude.

Munich

$$p_1 = 1$$

$$p_2 = 10/s_{km}$$

$$p_3 = 100/s^2_{km}$$

$$p_4 = 100/(56 + 0.20s^2_{km})$$

TABLE 1. Summary of Mean Errors

Distance, km	Frankfurt, cm	Hanover, cm	Münster, cm	Munich, cm	Average, cm
0.5- 2	± 4.0	± 3.0	± 2.0	...	± 3.0
2- 5	7.6	6.3	7.0
5-10	1.4	3.5	...	± 4.6	3.2
10-20	8.3	3.6	...	1.7	4.4
20-40	8.0	3.7	...	5.6	5.8
>40	6.4	7.9	7.2

From all measurements and for all distances, ± 5 cm

The adjustments with different weights gave results with only minor differences. The adjustments with the weight $p_1 = 1$ showed the following mean errors of unit weight: Hanover: $m = \pm 5.9$ cm; Munich: $m_0 = \pm 3.2$ cm. In the city network of Hanover the average difference between the lines computed from coordinates and from the Tellurometer lines after the adjustment amounted to ± 5 cm.

RESULTS

In Table 1 the results of the Tellurometer measurements are listed according to the different areas and distance groups. The indicated mean errors m are computed from the probable deviations as well as from the differences of observations, and they represent only the internal accuracy. The differences between the lines computed from coordinates, listed in international meters, and reduced Tellurometer distances contain the errors of the electronic distance measurements as well as the errors of triangulation lines. These deviations are still being discussed.

The average value shown in Table 1 for the internal accuracy of a line measurement is

very favorable, especially considering that a part of the error resulting from various influences of topographic and meteorologic conditions prevailing during our measurements is contained therein.

The external accuracy can be given reliably only after a comparison with the results of the base expansion Munich.

CONCLUSION

The Tellurometer equipment was prepared in the autumn of 1958 for use in the Arctic, and test measurements have been carried out in an air-conditioned room down to a temperature of -40° C as well as in practical measurements on the Zugspitzplatt (High Alps) at a height of 2600 meters and in the valley of the Isar river (near Munich) at low temperatures and above a solid coverage of snow. The instruments are now in Greenland with the International Glaciological Expedition, where a chain of quadrilaterals with a length of more than 800 km across the inland ice is to be measured. A later discussion of these results should be of special interest.

Remarks concerning Current Use of the Tellurometer

G. COETS

*Institut Géographique de the Belgian Congo
Elisabethville, B. C.*

Introduction. The Institut Géographique du Congo Belge (IGCB) has viewed with interest the development of instruments for measuring lengths, such as the Geodimeter and the Tellurometer, based on the speed of wave propagation. Once this development succeeded in producing instruments, simple, sturdy, light in structure, and easy to handle, their use was bound to improve appreciably the productivity of the individual operator. There would logically result a decrease in the cost of operation as well as a saving of time, both of which are particularly important in underdeveloped countries.

In mid 1958 this stage of development seemed to have been reached, and the IGCB has since then purchased one Geodimeter type NASM 3, and three masters and five remotes of the Tellurometer type M/RA 1.

No systematic trials with the Geodimeter have, as yet, been attempted, but a systematic program of trials with the Tellurometer is in progress. It should be noted that this instrument has been studied, not from the standpoint of maximum efficiency, but with a view to determining what may be expected under normal conditions in an equatorial region where immediate and practical results for the development of the country are more important than the more remote scientific aspects.

PRECISION OF THE TELLUREMETER IN THE FIELD

Nomenclature

In what follows the word 'error,' unless otherwise specified, will be used in the sense of mean square error, and 'relative error' in the sense of relative m.s.e. The corresponding symbols are ϵ and ρ . Relative errors, being of the form dl/l and dimensionless, can be expressed either as a fraction (e.g., 1/100,000 or 10^{-6}) or in seconds of arc ($1'' = 1/206,265$). Statistical values of ϵ and ρ , obtained from an adjustment,

are designated η and τ . These values result from the use of $\sqrt{\sum v^2/n_c}$, where n_c represents the number of conditions.

Preliminary Considerations

The problem of the precision of measurement with the Tellurometer is complex, considering all the various factors that can affect it.

We note first of all that the value used for the speed of light introduces a slight but systematic lack of precision.

One error, which is both accidental and systematic, arises from the definition of the principal frequency used (called frequency Λ). It is affected by the aging of the crystal, by the validity of the calibration curve of the crystal, and by the errors in reading the state (temperature) of the crystal from a voltmeter. These errors are of no great consequence for the present use, but for control of taped base lines, or for measuring new bases, an attempt should be made to eliminate them by bracketing the measurements between two calibrations, to make certain that no breaks have occurred during the course of the measurements.

Another source of error is introduced by atmospheric conditions, from the fact that the average value is used based on measurements at the two ends of the line; also, errors are inherent in the formulas themselves.

A further unpredictable error results from adopting a mean of x measurements to eliminate 'swing', a phenomenon whose causes, though known, are difficult to assess.

Reobservation of Base Lines

Three projects have been carried out; we have remeasured the bases of Kitona (6 km) and Kitomesa (6 km) as well as a side of the Limbe Base expansion (13 km). These experiments are of interest both on account of the measurement

selves and because of the results of the analysis of the measurements.

Remeasurement of Kilona base. This base, approximately 6300 meters long, was measured by Invar wire, after necessary terrain preparation, in 8 days by a team of 8 men. Remeasurement by Tellurometer was accomplished by 2 operators in 3 days' work, each consisting of a series of 12 measurements, the series beginning with the frequencies 1, 4, 7, 10 ... and progressing by $\frac{1}{4}$. The purpose for this multiplicity of measurements was control. The data follow:

First day, from 16 to 18 hours, strong wind.
Second day, from 9 to 10 hours, master and remote interchanged; weather calm and humid.
Third day, from 8h30 to 10, master and remote on first day; weather calm and humid.

Length of base by Invar wire: 6269.696 meters.
In what follows, only the fraction of a meter above 6269 will be cited, expressed in millimeters.

Mean lengths for the 3 days: 715, 690, and 679 mm.

Value of η , for each day, for a series of 12 measurements, obtained from the dispersion with respect to the mean for the day, by $\sqrt{\Sigma v^2/5}$: 24, and 18 mm.

Mean length for the set of 3 days: 679 mm.
for a series, by $\sqrt{\Sigma v^2/17}$: 47 mm.

of the mean, obtained from $\eta/\sqrt{18}$: 11 mm.

The theoretical error, according to the manufacturer, for a series of 12 measurements, obtained by quadratic combination of 50 mm with 1/300,000 of the distance, is 54 mm.

It can be seen that the *apparent* value of η agrees closely with the theoretical value of 54 mm, and that the error obtained for the mean, $696 - 679 = 17$ mm, is on the order of the statistical value of 11 mm.

Furthermore, using the concept of deviation from the mean, defining the zone within which a per cent of the measurements should be enclosed, it is found that this zone defined by $\pm (2/3)(54)$, i.e. 660 to 732, contains 10 of 18 measurements.

For the zone twice this width, corresponding theoretically to 82 per cent of the measurements, extends from 624 to 768 and is found to contain 15 measurements, or 84 per cent.

There remains an examination of the daily results. Since there are 6 series per day we may expect for the daily mean a theoretical error of $54/\sqrt{6} = 22$ mm, or a probable error of 15 mm.

The 3 days yield differences of 19, 6, and 63 mm (with respect to the Invar value). The last difference may seem excessive, but it must be borne in mind that these experiments are not carried out under identical conditions. Just as it is illusory to assume that, with a 1-second theodolite, 100 readings will reduce the error to 1/10 of a second, so it is illusory to believe that increasing the number of series will improve indefinitely the precision of the mean.

In conclusion it is admitted that, under normal conditions, the theoretical error of a series as stated by the manufacturer is correct, and that for precise work it is of interest to increase the number of series, if only to give play to varying atmospheric conditions, but that the effect on the error will be slight.

The base could therefore have been remeasured in 6 series on 1 day and still conform to the theoretical precision; working time would have been of the order of 2 hours.

Remeasurement of the Kitomesa base. The base, approximately 6400 meters long, was measured by Invar wire, after terrain preparation, in 8 days by a crew of 6. Remeasurement with Tellurometer was accomplished in 2 days by 2 operators, measurements being spread over the period from morning to evening. The site being unfavorable—flat, stony terrain—it was decided to measure the base not only directly but in two legs as well.

Length of base by wire: 6366.366 meters.

Difference: 65 mm (with respect to Invar value).

η for 1 series from $\sqrt{\Sigma v^2/7} = 120$ mm.

Manufacturer's theoretical error = 54 mm.

η_m of the mean, $120/\sqrt{8} = 43$ mm.

Dispersion with respect to the mean being relatively wide we give the deviations with respect to the mean, in millimeters: -24, +86, -234, +86, +46, +146, -74, -14.

Undoubtedly this dispersion is due to the unfavorable site; however, agreement with the errors 65, 54, and 43 is good. The results seem to confirm the conclusions of the preceding section.

The worst series yielded a value of 6366.067 with an error of 299 mm, which is slightly more than 5 times the theoretical error but only 2.5 times the statistical error.

The base was then divided into two sections and each was measured in 4 series. The resulting length was 6366.450 meters, differing by 84 mm, with a theoretical error of about 70 mm.

The weighted mean of all measurements was 6366.330 with a difference of only 36 mm.

The opportunity presented itself to see whether a constant or so-called zero correction could not be deduced from this experiment. If x is the constant correction, D the measured length, and D_1, D_2 its component lengths, then:

$$D + x = D_1 + D_2 + 2x$$

$$x = D - D_1 - D_2$$

In the present instance $x = -149$ mm, which cannot be considered.

Measurement of a side in the Limete baseline expansion. The side measures 13,400 meters and has a m.s.e. deduced from the adjustment of 1/175,000, or 7.7 cm. This differs from the length obtained with the Tellurometer in 4 series of 12 measurements by 17 cm, which is quite normal. Combining the theoretical error of 6.7 cm with the 7.7-cm error gives 10.2 cm, which shows the error found, 17 cm, to be quite acceptable.

Measurement of a Complex Figure at Matadi

A figure consisting of a central quadrilateral and its two diagonals was completely observed with Tellurometer and theodolite. The sides were measured in 4 series of 12 measurements, and the angles with a Wild T2 in 8 readings.

Considering only the lengths this figure requires three equations of condition. The adjustment was made assuming the relative errors of the lengths to be constant. The unknowns of the form dl/l were expressed in seconds of arc; the unit of variation is therefore 1/206,265.

The adjustment yields from $\sqrt{\Sigma v^2/n_e}$ a statistical value τ for the relative error ρ of a measurement. We find $\tau = 1.12''$ or 1/184,000. Considering the angles only, the adjustment yields the value $1.56''$ for the statistical value η of the error ϵ in the measurement of an angle. The combined adjustment could have been made

by introducing weights according to the ratio $(1.56/1.12)^2 = 2$; but, since some of the series were somewhat unfavorable for measuring angles, no special precautions had been taken, it was decided to adjust with equal weights, with the resulting values $\eta = \tau = 1.86'' = 1/110,000$. The combined adjustment indicates a lack of homogeneity between the measurement of angles and lengths. In view of the fact that the sides varied between 9 and 26 km, with corresponding variation in the theoretical value of the errors from 1/150,000 to 1/250,000, appears likely from the results that the introduction of weights would have given better agreement. This question must, for lack of time, be deferred.

The residual error in the adjusted values obtained from $\eta' = \eta\sqrt{(n - n_e)/n}$ and is practically $1''$, or 1/200,000. We may conclude that, in the current operations, the measurements of lengths and angles yield comparable precision.

Results of Some Check Measurements

(a) Control of a 10-km side at Kikwit: 7 series of 12 measurements; statistical error of 5 cm for a series, 5 cm.

(b) Control for a 35-km side of triangulation: 7 series of 12 measurements; value of η , 14 cm.

(c) Control for 5-km side at Luluabourg: 7 series of measurements; value of η , 6 cm.

Comparison with the theoretical error, obtained by combining 5 cm with 1/300,000 of distance, yields the following:

	Statistical Error,	Theoretical Error
	cm	cm
5 km	6.4	5.3
10 km	5.2	6
35 km	14.3	12.7

The agreement with the statistical error derived from $\sqrt{\Sigma v^2/(n - 1)}$ is remarkable.

Measurement of a very Short Distance

For experimental purposes the 100-meter base at Kasoma, established with Invar wire and used to calibrate steel tapes, was remeasured with the Tellurometer by difference of the lengths of approximately 200 and 300 meters in 4 series of different set-ups. The experiment yielded varying results according to the

encies used, errors going up to 8 cm, the results being obtained from use of the frequencies, chosen in the lower range (5-10) of the instrument.

This confirms the manufacturer's statement that the Tellurometer was not built for measuring short distances. Since modern practice in triangulating with Invar stadia yields a precision of the order of $1/6000$, it seems practical to consider $18 \times 6000 = 300$ meters as the lower limit of usefulness for the Tellurometer. Moreover, if a measurement by difference with the Tellurometer takes about half an hour, it will require much more time to measure 300 meters with the stadia (2 sections of 150 meters, the theodolite set up in the center of each station).

Conclusions

The manufacturer's claim for precision seems justified under normal conditions and without special preparation. For high-precision work, however, the Geodimeter seems to us superior to the Tellurometer for the following reasons: (a) the crystal operates at constant temperature; (b) swinging is absent; (c) there is less effect from errors in the observation of atmospheric conditions.

On the other hand, the Tellurometer is, at present, more practical for all other operations because of its light weight, ease in handling, and set-up, and, of particular importance in Africa, because it affords the possibility of day-long operation (mosquitoes at night) and in all weather.

MODIFICATIONS IN METHODS; ECONOMICS OF THE TELLUROMETER

As stated before, increased production from the operator rather than increase in the precision of the measurements is our aim in the Belgian Congo in the use of the Tellurometer. Work is in progress to determine its utilization in various phases of surveying. Definite conclusions having been obtained in a first type of operation (below), the equipment for this type of operation will in future include Tellurometers (master, 2 remotes). Studies on the other phases of work are incomplete, and no final conclusion can as yet be reached.

Topographic surveys for normal intermediate

scale mapping. Photogrammetric surveying requires the identification and coordination of a number of prominent geographical features in the terrain and the corresponding stereogram. In general, 6 points must be located planimetrically and 6 in elevation, or sometimes in combination when they meet the required conditions.

The area covered by a stereogram varies, but, the mean scale for the photographic coverage of the country being around $1/40,000$ and the size of the film 18 by 18 cm, it is usually 6 by 3 km. The precision of the topographic survey is governed by the optimum return from the photographs. The field party assumes the existence of an ordinary triangulation network (sides approximately 15 km) starting from which one or more points in the stereogram or its immediate vicinity can be determined. The group of 6 points is tied in to these by the various known methods of measurement of angles and sides.

The appearance of the Tellurometer simplifies the problem considerably by decreasing the number of check points needed to avoid gross errors, by eliminating the conditions relative to strength of figures, and especially by permitting the measurement of relatively long distances which would be prohibitively time-consuming with classical methods. Thus, for example, the central point of the stereogram instead of needing to be fixed and leveled from 3 visible points can be determined from a single known side from 3 elements, angles or lengths. From this central point it will in general be possible to fix others merely by measuring an angle and a length with the Tellurometer; at times a traverse may be necessary in addition, but its length will always be very much reduced.

In short, the gain in time for an operator with some knowledge of the terrain can be important especially in the tropics where access and intervisibility are often difficult. Each new station must be recovered, occupied, and marked. Every method that permits the reduction in the number of intermediate stations represents a saving in time.

For example it is estimated that a stereogram occupies a crew of 2 Europeans for 2 months. Since the introduction of the Tellurometer this time has dropped to 2 weeks or even 10 days. Moreover, native topographers have been used

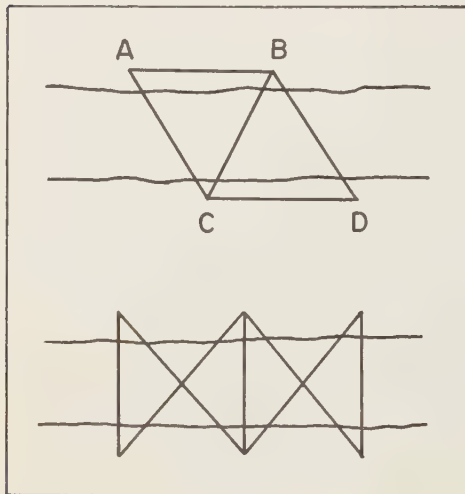


Fig. 1

as operators of remotes and even masters, revealing in practice more talent than in operating a theodolite.

Local surveys with or without photogrammetric compilation. This is generally a mixed local survey, triangulation and traverse, with a density of $1/2$ to 1 mark per km^2 , established in a city of some importance and its immediate vicinity. It is from these marks that a surveying party proceeds, for compilation purposes, by the classical traverse method. One experiment is in progress. The gain in time is being tested in replacing the slow ordinary traverse method by Tellurometer polygonation wherever possible.

River triangulation. In tropical regions triangulation along rivers is made difficult by the

sides along the banks. Sides such as AB and CD (Fig. 1) even along the concave side, are often difficult to observe without arduous clearing of brush and trees. This often leads to very short lengths with a corresponding loss of precision and time.

An experiment will shortly be carried out along a 120-km stretch of the Congo River with 2 Tellurometer traverses, crossed and braced (as shown in the figure).

It will therefore not be necessary to measure along the banks unless it is convenient, and the lengths will be limited only by intervisibility across the river, which often permits considerable lengths.

Ground equipment for photogrammetric compilation on small scale in unsurveyed areas. It is proposed to establish astronomic points, 8 per square degree, and to compile by the template method. The typical working party includes European, sometimes 2 (shortage of native recorders); it covers on the average 4 degrees square per year. Replacement of this procedure by Tellurometer traverse in the portions where the terrain is suitable is contemplated, in the hope of increasing the production of the parties. Moreover, the use of trilateration in certain cases is under consideration.

Production of additional data by leveling parties in unsurveyed areas. Concurrently with their normal operations it is contemplated to have leveling parties, operating in untriangulated areas, run a Tellurometer traverse. The relatively slow progress of these parties (80 km per month) should enable them without much loss of time to obtain in addition horizontal control from such a traverse.

Tellurometer Trilateration in Arabia

HANS MEIER

*Aero Service Corporation
Philadelphia, Pennsylvania*

Since the introduction of electronic distance-measuring equipment for medium-length distances (1 to 30 miles), much has been published about the principles and accuracy of the various instruments. The author wishes to relate some practical experience with the Tellurometer during a trilateration survey in the Arabian desert. In 1958, Aero Service Corporation was given the problem of determining the coordinate position of four offshore navigation beacons near Ras Tanura in Saudi Arabia for the Arabian American Oil Company.

The horizontal control points that were to be used as the base of the survey could not be recovered with certainty on the ground and so were of little value. It was, therefore, necessary to extend the survey to assure positive recovery of the previously established control points. To do the job by triangulation appeared difficult, owing to the adverse weather conditions prevailing in the area. Sandstorms impair visibility, and costly tower construction would have been necessary at various stations to assure line of sight. These considerations and the fear of large lateral refraction over the desert led to the conclusion that the problem could be solved economically only by Tellurometer trilateration.

Figure 1 shows the trilateration network which was measured by a Tellurometer party of 4 men (2 surveyors and 2 Arab helpers) in 10 days. Fifty-nine distances were measured, varying in length from 3.5 to 58.2 km. Seven of them were over water, these distances usually being looked upon with suspicion. Any flat, smooth surface over which the signal travels is known to reflect the ground wave and will, therefore, interfere with the direct wave signal, resulting in a fuzzy scope presentation and generally an unreliable distance measurement.

None of these phenomena, however, were experienced with the *long* over-water measurements that were made at night for logistic convenience rather than on purpose. The short over-water distances measured during the daytime, however, showed slightly fuzzy scope presentation and erratic distance readings.

All 4 distance measurements to beacon 4 deserve special mention, because the remote station at the beacon site could not be seen from the master station locations on shore. The chord of the ray path actually cut 14 to 69 ft below the water surface. No trouble was experienced in measuring these lines. In contrast to this rather unusual phenomenon, it was impossible to measure the much shorter distance from beacon 6 to beacon 4. It was not even possible to get voice communication over the beam, and a messenger had to be sent by boat to relay messages. The reason for this strange occurrence is a mystery and would certainly be worth investigating.

During the whole operation, the survey party was plagued by equipment failures. Cable contacts broke too easily, and sand penetrated into the instrument. The sandstorms blowing over the desert made the vertical-angle measurements difficult and slowed the operation, since visibility was very poor. Trying to solve this survey problem by triangulation would have been uneconomical. Triangulation as well as traverse surveys, both depending on horizontal-angle measurements, are not suited for desert areas, because of the unfavorable influence of prevailing weather conditions on the measurements. Trilateration measurements are not restricted by these conditions and can be made under otherwise impossible circumstances. The Tellurometer party experienced the handicap of angular measurements vividly when considerable time was wasted waiting for the at-

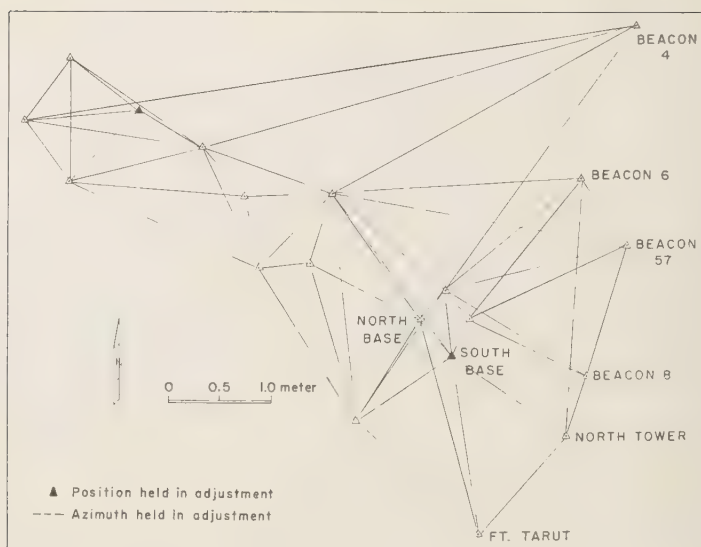


Fig. 1. Ras Tanura Tellurometer trilateration.



Fig. 2. Probable point errors in meters.

mosphere to clear in order to make the vertical-angle readings, necessary for the slope correction of the Tellurometer measurement.

Anyone who has had the frustrating experience of sighting on a damaged triangulation

signal can readily appreciate the great advantage of trilateration measurements. The remote station is attended and therefore cannot be destroyed by animals or local inhabitants while it is in operation.

It is our contention that the triangulation method would have been impossible for solving the survey problem, both from a technical and an economical point of view.

Figure 2 shows the magnitude of the probable point errors. In the adjustment, two stations and one azimuth were held fixed. The average probable error of all points is ± 0.33 meter. The average probable position error of four beacons is ± 0.43 meter (± 1.3 ft), which shows clearly that the over-water distances are not of the same quality as the over-land measurements. The trilateration network was planned to satisfy specifications which allowed a probable position error of ± 0.76 meter (± 2.5 ft) for the beacon sites. This requirement was fulfilled as shown by the above error figures. Table 1 gives a summary of the probable position error of each station.

This survey proved the advantage of the trilateration method over triangulation in an area with poor visibility conditions. Using a laser and two remote stations would have expedited the operation considerably, since most of the unproductive time was spent in traveling from point to point.

TABLE 1. Probable Errors of Observations

The probable error of a single observation is ± 0.207 meter.

Station	Probable Error of Each Position	
	Latitude, seconds	Longitude, seconds
1	± 0.013	± 0.006
2	0.005	0.007
3	0.009	0.009
4	0.009	0.006
5	0.009	0.006
6	0.008	0.008
7	0.008	0.008
8	0.010	0.008
9	0.006	0.007
10	0.004	0.006
11	0.006	0.006
12	0.003	0.003
13	0.005	0.015
14	0.010	0.008
15	0.013	0.010
16	0.010	0.009
17	0.010	0.010
18	0.012	0.007

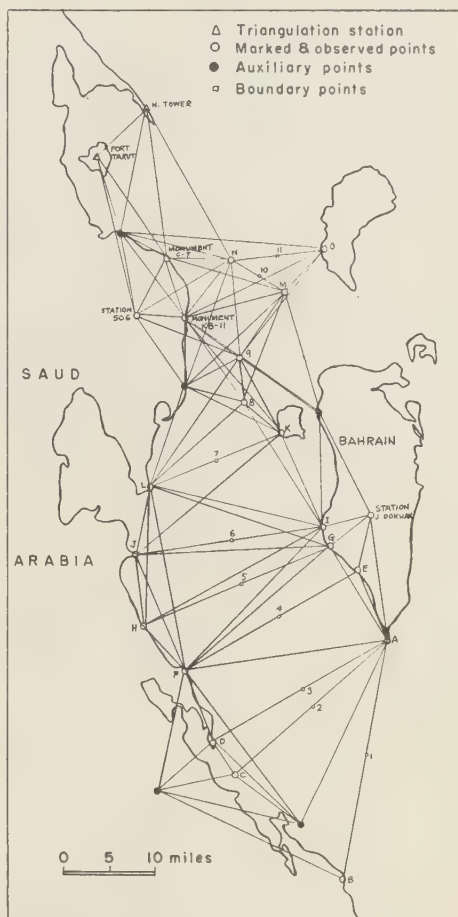


Fig. 3. Proposed trilateration scheme to establish the Saudi Arabian-Bahrain boundary.

Another Tellurometer trilateration is in progress at the present time in the same area. Its basic concept is a direct outgrowth of the experience gained during the Ras Tanura survey, which showed that over-water distances can be measured without line-of-sight connection between the two Tellurometer stations.

Aero Service is now engaged in a sea boundary survey between Saudi Arabia and Bahrain. Figure 3 shows the trilateration network and the boundary points. The boundary position will be determined by halving the distance be-

tween opposite shore markers. No tower construction is anticipated. This Tellurometer party is equipped with one master and two remote stations in order to work more efficiently. As of today, half of the project is completed.

The difficulties encountered so far are not of a technical nature but are due to vehicle trouble. More and more we realize that adequate means of transportation and logistic support are the deciding factors on a Tellurometer trilateration project, since the time required to make a distance measurement is negligible compared with the time spent to get to the station site in difficult terrain.

It is well known that all Tellurometer measurements are subject to ambient meteorological conditions which change the velocity of the emitted signals inversely to the index of refraction of the transversed medium. Pressure and dry- and wet-bulb temperature measurements are usually made at both stations, and the average index of refraction is used to reduce the Tellurometer readings. This procedure gives satisfactory results. Discrete measurements along the signal path, however, would increase the distance-measuring accuracy, since a change in the sixth decimal place of the index of refraction affects the centimeters, that is, the inherent measuring accuracy of the instrument.

In order to facilitate the index of refraction computations, Aero Service uses a method developed by Captain Carl Aslakson. The basic formula for the index of refraction is a function of pressure and dry- and wet-bulb temperature. It can be separated into three parts, which are

separately functions of wet- and dry-bulb temperatures and pressure. Two nomograms were constructed which are entered with the three measured parameters, extracting two values L and N_s , which are added to obtain the index of refraction.

Over-water measurements without line of sight between the two stations are considered as geodesics on the spheroid, and no slope or sea-level corrections are applied. The least squares adjustment of the Tellurometer trilateration network is made by the variation of coordinates methods used in Shoran trilateration adjustments.

When planning trilateration networks, it is necessary to keep in mind that more lines must be measured than in a triangulation scheme. The same strength of figure is to be maintained. Trilateration networks are, therefore, more elaborate and not as symmetrical as triangulation chains of quadrilaterals. In order to detect any possible constant measuring error K , it is advisable to insert 'sliver triangles,' which are long thin triangles, in the trilateration network. This configuration is foreign to triangulation networks. If there are enough sliver triangles in a network, and if a constant measuring error is actually present in all distance measurements, the value of K can be determined by inserting K as an unknown in the observation equations.

This description of Aero's experience with Tellurometer trilateration in Arabia demonstrates, I hope, the practical significance of electronic trilateration surveys.

Tellurometer Operations in Topographic Mapping

Abstract

JULIUS L. SPEERT

*U. S. Geological Survey
Washington, D. C.*

The U. S. Geological Survey uses the Tellurometer principally to increase the speed and reduce the costs of control surveys for topographic mapping. Increased accuracy is a welcome by-product. In this application, it largely replaces triangulation and transit-and-tape traverses. Most of our Tellurometer work consists of traverse, with angles measured by a 1-second optical theodolite. Second-order accuracies are obtained at lower costs than for previous third-order work. All traverses are tied to the basic geodetic net. Where geodetic azimuth ties are readily available, azimuths are obtained from Polaris observations. Our basic traverse observations are supplemented, where appropriate, by a small amount of trilateration, numerical side shots or radial measurements from strategically located control stations, and triangulation intersections of conspicuous points.

The Tellurometer is also effective for testing the accuracy of finished maps.

This report presents some statistical data on quantities, accuracies, and costs, developed during about a year of operations of four field parties.

Data-reduction problems included development of a visual aid for interpreting Tellurometer coarse readings and a plastic computer for obtaining the semivelocity of the radio wave in the atmosphere as a function of wet- and dry-bulb temperatures and barometric altitudes. Vertical-angle measurements are used to reduce Tellurometer distances to horizontal and to sea level, and to compute elevations. The resulting geographic positions and elevations are adjusted by simple proration or by least squares, as appropriate.

Geodetic Base Lines

W. A. HEISKANEN

*Institute of Geodesy, Photogrammetry and Cartography
Ohio State University
Columbus, Ohio*

We need in geodesy different world systems to meet the present high requirements of this atomic and space age. A world gravimetric system almost exists now, thanks to the efforts of the geodesists of various countries, particularly Dr. G. P. Woollard's group at the University of Wisconsin. We are also very close to the end goal of geodesy: to determine in detail the shape of the geoid and to establish the world geodetic system, mostly through world-wide gravity studies.

We have been able to compute the shape of the geoid with relatively high accuracy. These methods, however, cannot give the size of the earth. For that purpose we need the arc measuring method, which requires accurate triangulation chains and good astronomic observations. By the angle measurements of the triangulation chain we get relatively accurately the shape of the chain but not its size. To get the scale we must have a geodetic yardstick, i.e., field base lines measured with high accuracy. In world-wide computations it is necessary to have a uniform scale for the yardsticks of different continents and countries. What we need is an exact geodetic world scale. This can be furnished only by geodetic standard base lines measured in different parts of the world.

All this not only has academic interest but is of basic significance for present-day geodesy. The studies carried out in various countries show that the existing scales are not yet sufficiently unified. For instance, the scale of the Baltic Geodetic Commission does not exactly agree with the scale used earlier in central Europe.

The required accuracy of the standard base lines depends first on the accuracy obtainable in the triangulation and in the measurement of the field base lines. The accuracy varies in dif-

ferent countries. According to the adjustment of the European triangulation in the U. S. Coast and Geodetic Survey, under the leadership of C. A. Whitten, the accuracy of the triangulation in Europe can be seen from the following figures: The average rms error of a direction was 0.51" in the countries of southwestern Europe, whereas, in the countries of northern Europe, it was 0.18" in Estonia and Latvia, 0.17" in Finland, 0.26" in Denmark, 0.45" in Norway, and 0.27" in Sweden. This large difference in the errors may be due to the fact that the triangulations were carried out much earlier in southwestern Europe than in the northern countries; Finland, Estonia and Latvia started their triangulations of first order only a few years ago, when more accurate instruments and better observation methods were available. This statistic shows that the average rms error of the direction can be even less than 0.2", or in linear units 1:1,000,000. Almost exactly this accuracy has been obtained in the Baltic geodetic ring measured and computed between 1924 and 1946. The closure error of this approximately 2500-km-long ring is only 2.5 meters or exactly 1:1,000,000.

The accuracy of the field base lines varies from 1:2,000,000 to 1:4,000,000. Consequently, we must have higher relative accuracy in the measurement of the standard base lines.

Nowadays Invar wires 24, 25, or 50 meters long are used in field base-line measurements almost exclusively. The wires must be calibrated under exactly the same conditions as will prevail in the field. The bar comparator of the length of the Invar wire will not yield the highest accuracy for the calibration because the conditions are not the same as in the field measurements. Moreover, the whole error of the comparison will enter in the end result.

need longer yardsticks or standard base lines which are at least 20 times as long as the Invar wires. Since, in general, we measure the standard base lines in the air from tripod to tripod, from pole to pole, the calibration of the tapes along the standard base line must be done in the same way, using exactly the same techniques, the same measuring method, and, if possible, the same men as well.

As far as I know, there exists only one such method, the light-interference method invented in 1923 by Y. Väisälä and put into practice by the Finnish Geodetic Institute. As the director in absentia—of this Institute, I am pleased to tell about our efforts to establish the world standard base-line system measured by the light-interference method. I shall not go into detail, because my colleague and friend, Professor Honkasalo, who has participated in all base-line measurements of this type will do so and can do it better.

The founder and first director of the Finnish Geodetic Institute, Professor I. Bonsdorff, suggested that Professor Väisälä construct a light-interference comparator for the Finnish Geodetic Institute. This comparator was exhibited at the Assembly of the IUGG in Stockholm in 1930. It was then mounted at the Finnish standard base line of Nummela, about 40 km northwest of Helsinki. This base line had earlier been measured many dozens of times with different Invar wires and was later on used for the standardization of many European Invar wires.

Even earlier, at Professor Bonsdorff's suggestion, the IAG had passed a resolution in which it was suggested that standard base lines should be established in as many countries as possible. This resolution produced no result. Hence, after the Nummela base-line measurement in 1947, made by Dr. Kukkamäki and Dr. Honkasalo, succeeded so well that the rms error was only $1 : 17,000,000$, the Finnish Geodetic Institute sent a circular to several countries on different continents recommending that standard base lines should be measured by means of the Väisälä light-interference method in different countries, at least one on each continent. We received a positive answer from Argentina. The negotiations were started, and in 1953 Dr. Kukkamäki and Dr. Honkasalo, at

the invitation of the Government of Argentina, measured the Buenos Aires base line of 480 meters with an accuracy of $1 : 9,000,000$.

The question about the standard base lines was also in the program at the General Assembly of the IUGG in Rome in 1954. This Assembly passed the resolution:

The International Union of Geodesy and Geophysics resolves that these member countries adopt the following program in so far as possible:

(3) Establish a standard base line in each country using the Väisälä method (or similar apparatus) for assuring a uniform scale in all networks and for calibrating Invar tapes and Geodimeters.

Similarly, the UN Regional Cartographic Conference for Asia and the Far East in Mussoorie in February 1955 adopted this resolution

The United Nations Regional Cartographic Conference, in view of the remarkably accurate results that are obtainable with the Väisälä Comparator of the Finnish Geodetic Institute, which has been calibrated against the international prototype of the International Bureau of Weights and Measures, and in view of the difficulties and uncertainties associated with the physical standards of length for calibration purposes, recommends to the Governments of the Asian countries that a few standard base lines in this region should be established by the Väisälä method for assuring a uniform scale in all networks and for calibrating Invar tapes and other equipment.

Soon after that negotiations were started with the geodesists of Holland, and in the fall of 1957 Kukkamäki and Honkasalo were invited to measure the Loenermark base line in the Netherlands. Its distance is 576 meters, and the rms error was about $1 : 11,000,000$. The Munich standard base line in Germany was measured in 1958 in two parts, the northern part in June and the southern in October. Dr. Honkasalo participated in the measurements of the southern part; the accuracy was of the order of $1 : 20,000,000$.

As you will hear in the next address, the Väisälä light-interference method is very simple in principle. We make consecutive multiplication of the standard meters so that we get first the length 6 meters, then 24, 72, 260, 432, and 864 meters. The error of these multiplications cannot be large because the multiplication is based on the light-interference figures. If the larger distances are not exactly an integral

number—in our case, 6, 4, 3, or 2 times larger than the smaller—we cannot see any interference figures. The weather conditions seem not to have too much effect if measurements are made on cloudy or rainy nights when the temperature along the base line is almost constant.

We have now the standard base lines of Nummela, of Buenos Aires, of Loenermark, the Netherlands, and of Munich, Germany, measured by the Väisälä interference method as the

Assembly of the IUGG in Rome suggested. It is the hope of geodesists that this method will be used also in the other countries. In fact negotiations have already been started with some other European countries. We also hope particularly that the United States and Canada will be interested to have such a standard base line. I myself have tried to get such a line measured in Columbus, but the finances of the Ohio State University have so far not permitted it.

Measurement of Standard Base Line with the Väisälä Light-Interference Comparator

T. HONKASALO

*Finnish Geodetic Institute
Helsinki, Finland*

The Väisälä comparator. Dr. Y. Väisälä presented his new method of measuring long distances with light interference 36 years ago [Väisälä, 1923], and the instrument was fully developed by 1929. Measurements of up to 100 meters succeeded with an accuracy of 10^{-7} . The principle of the Väisälä comparator can be seen from Figure 1. White light from the source J is made parallel by a collimator lens L . The beam is divided at mirror M_2 into two. While one beam passes through the hole of mirror M_2 to the middle point of mirror M_1 , where it is reflected back through the other hole in mirror M_2 , the other beam is reflected a certain number of times between the mirrors M_1 and M_2 . The two light beams meet in the focal plane of the observation telescope T . If the paths traveled by the two beams are equal, the diffraction interference lines appear in the eyepiece J . If monochromatic light is used the interference can be seen though the difference in length of the paths traveled by the beams is many millimeters, but in white light the lines appear when the difference is but a few light wavelengths, as a result of the different diffraction of different colors. On the basis of symmetry of the diffraction interference lines the

equality of the light paths can be observed to an accuracy of one light wave.

To obviate the necessity of adjusting the mirrors in the right position to an accuracy of 1 micron, a compensator of two plane-parallel glass plates P is used. By turning one of these glass plates, the optical distance of the beam passing through it is extended and the interference lines can thus be centered symmetrically onto the image. The compensator can be used as a measuring device; by reading the angle of rotation the difference in the light-beam paths can be computed. An ocular prism is used to find the correct position for the glass compensators. With it the distance error can be 20 microns because the interference can be seen in any part of the colored picture, and there are more interference lines because the light in every part of the spectrum is nearly monochromatic. When the interference has once been seen, the mirrors are moved to a position in which the angles of rotation of the compensators are small. Small angles are also used by achromatic compensators, which are accurate up to very great angles, because the reading accuracy with small angles is greater.

Quartz gage. The first mirror distance is

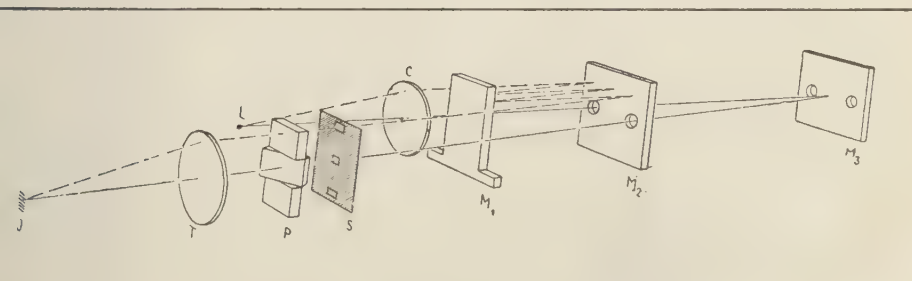


Fig. 1. Diagram of Väisälä light-interference comparator.

not measured directly by light interference in the field, but with a quartz meter, the length of which is compared in the laboratory with the wavelength of light.

The quartz gage is made of a fused-silica tube with spherical polished quartz-glass endings. At one end the radius of curvature is 1 meter, to make it possible to bring the gage into optical contact with the mirror surface, and at the other end 5 meters, to make it easy to measure the small gap between the gage end and the mirror surface by means of Newton fringes.

The quartz meters are good standards because of the small dilation constant, 0.4×10^{-6} , and their stability. Kukkamäki's thorough study and the intercomparisons of these meters made during the past 25 years show the excellence of the stability. Since the intercomparison of these meters in a special comparator is possible to an accuracy of ± 0.004 micron, these gages form a special system of their own. This system is compared with the international meter by comparing one of them (no. XV) with the Finnish meter prototype 5 [Kukkamäki, 1933] and two others, I and VIII, with the German meter prototype in Charlottenburg. In 1954 quartz meters VIII and XI were compared with the wavelength of light (green line of mercury 198) at the Bureau International des Poids et Mesures in Sèvres. Thus, when in the near future the international meter is defined in wavelengths of light, the base lines measured by the Väisälä comparator will already have been measured in international meters.

Choosing the site for the interference standard base. The International Union of Geodesy and Geophysics has recommended the establishment of international standard base lines measured by light interference. Accurate measurement by light interference is such a laborious task that it cannot be done every time the standard base line is used. It is therefore very important to choose the ground so that the base line remains stable—a very difficult problem. Experience has shown that even bedrock cannot be considered stable when distances are to be measured with a high degree of accuracy. This is easy to understand when we consider that the rock is always split up into blocks and that stone has a relatively high temperature

dilation coefficient. Thus, the direction in which the mark moves depends on the corner of the block in which the bolt is fastened. Clay soil is most unsuitable. The marks shift when it rains, and in winter because of frost. The European interference standard base lines in Finland, in the Netherlands, and in Germany are on gravel soil. In Finland there is about 25 meters of moraine gravel under the standard base line. This base line has been measured with light interference four times, and the stability has proved very good. There are changes, it is true, but they have been so small that only accurate light-interference measurements have been able to ascertain them.

The terrain must be fairly even, because light travels in a straight line. The light should pass about 1 meter above the ground. At the height of the light beam many thermometers are suspended along the base line to control the differences in the refraction index. Depressions in the ground do not cause difficulties, provided they are not so large as to hamper the reading of the thermometers. On the other hand even a very small hill causes considerable disturbance, because the air temperature usually changes rapidly near the ground. It is advantageous, but not imperative, to build the base line in a forest, where differences in air temperature are smaller than in open terrain. The standard base lines, in Argentina and in the Netherlands, for example, are situated in open places.

The length of a standard base line must be sufficiently great to avoid the effect of any possible small changes caused by soil temperature variation or other movements in the ground on the Invar wire measurements. The standard base line in Finland is 864 meters, that in Argentina 480, in the Netherlands 576, and in Germany 864. Longer base lines than these are difficult to measure in one set.

Markers. With an interference comparator it is possible to measure the distance between the mirror surfaces with a high degree of accuracy, but to preserve the distance measured it must be transferred to underground markers. To ensure stability these markers must be fastened to massive blocks, perhaps concrete blocks 2 meters square or larger. If there are three bolts, one at either end and one in the

idle, it is possible to control the stability of the base line with Invar wire measurements. Because by usual measurements the difference between the two parts contains no systematic errors. The best possible form for a marker is a stainless-steel bolt 20 mm thick with a 3-mm well-made round hole on the top. The hole can be projected up to the Invar wire easily and accurately. A line on the top of the bolt is difficult to sight optically because of unsymmetrical light, and it can easily be imaged.

Pillars for the interference comparator. The interference comparator makes it possible to multiply a known distance by a small integral number. We start from 1 meter, multiply it two stages up to the length of an Invar wire, and then to the length of the base line. For example, on the Finnish base we have used multiplications $1 \times 6 \times 4 \times 3 \times 3 \times 2 \times 2$. Thus the pillars 0, 1, 6, 24, 72, 216, 432, and 864 are used in measurement, and in addition to these, one pillar for the interference telescope. The pillars are sunk 1 meter in the ground, and the height depends on the terrain so that the tops are in a straight line. The average height of the pillars is 1 meter.

The concrete pillars for the underground marks can be placed under the interference pillars if these are A-shaped; then the projecting of the measured mirror distance to the underground marks is easy.

Measurement with a Väisälä comparator. The supports for the comparator mirror are fastened on top of the concrete pillars. The mirrors are set perpendicular to the comparator line by the autocollimation method. The distances between the mirrors are measured with tape to an accuracy of ± 1 mm; the observation telescope and the lamp are installed on the telescope pillar. Accuracy of measuring depends on weather conditions. Short distances up to 100 meters can be measured in any weather, but for longer distances the refraction changes make the interference lines too movable. Longer distances are usually measured on a cloudy night. Rain does not harm, but fog makes observation impossible.

The small effect of refraction is corrected by suspending along the comparator thermometers shielded against radiation, which are

read by two assistant observers simultaneously with the optical interference observations. During all this time the observer follows the changing of the interference lines with compensators. When the mean of the compensator readings is taken, and the mean air temperature difference of the two light beams is computed, the refraction correction can be computed. A difference of 1°C means a correction of 10^{-6} . Since the first two pillars and the quartz meter are usually sheltered by a tent or roof there is a systematic temperature difference at the forepart of the comparator. To measure the air-temperature differences accurately at the shortest interferences we have constructed a special resistance thermometer. Air pressure and humidity have such a small influence on refraction that they can be ignored. If the comparator is inclined the systematic air-pressure difference between the ends of the comparator must be taken into consideration.

As the refraction changes from day to day the direction of the mirrors must be slightly corrected during the measuring period. Therefore the places of the mirrors on the pillars are always controlled by a 'plumbing device,' a special apparatus with a spirit level and a micrometer screw. On top of the pillars steel bars are firmly fastened perpendicular to the comparator line. The distance between the bar and the mirror surface is measured with the plumbing device.

When we measure, for instance, the 432-meter distance using mirrors at 0, 1, 6, 24, 72, 216, 432 meters, we begin from the longest interference, 0, 216, 432. When the thermometer and compensator readings are taken, we can compute the relation of the distances between the mirror surfaces 0, 216 and 0, 432. Now the mirror at 72 meters is placed on its support and the interference 0, 72, 216 is observed in the same way. The measuring is continued by interferences 0, 24, 72 and 0, 6, 24. At interference 0, 1, 6 the distance 0, 1 is measured with a quartz meter. The latter half of the observation series is made from the shorter interferences to the longest one. Every time a mirror is taken away from its support the position of the mirror on top of the pillar is checked with a plumbing device. One such series of observations takes about 2 or 3 hours.

On an interference standard base line we have usually made 4 to 6 series of observations on different days and in different weather conditions. The accuracy achieved is 10^{-7} of the measured distance. The time required for measuring a standard base line depends on the weather. If the concrete pillars are ready, we estimate that the installation of the comparator, the interference measurement, and the projection measurement from the mirrors to the underground marks takes about 4 weeks.

The measured interference standard base lines. The Nummela Standard Base Line has been measured four times, in 1947, 1952, 1955, and 1958. The first of these measurements has been published. When the comparison of the quartz meters made in 1954 with the wavelengths of light is taken into consideration, the following results are obtained:

	First Half	Total Line
1947	$432,095.02 \pm 0.03$ mm	$864,121.89 \pm 0.05$ mm
1952	$432,094.92 \pm 0.04$	$864,121.55 \pm 0.06$
1955	$432,094.78 \pm 0.05$	$864,121.49 \pm 0.08$
1958	$432,094.75 \pm 0.03$	$864,121.37 \pm 0.06$

The Buenos Aires standard base line in Ar-

gentina was measured in 1953. The result was

First Half	Total Line
$240,005.59 \pm 0.04$ mm	$480,001.78 \pm 0.05$ mm

The Loenermark base line in the Netherlands was measured in 1957. The result was:

First Half	Total Line
$288,051.63 \pm 0.03$ mm	$576,092.26 \pm 0.05$ mm

The Munich standard base line was measured in 1958 in two parts, the northern part in the end of June and beginning of July, the southern part in October. The author took part in the measurements of the southern part. The results were:

Northern Part	Southern Part
$432,031.66 \pm 0.08$ mm	$432,032.92 \pm 0.04$ mm

REFERENCES

- Kukkamäki, T. J., Untersuchungen über die Messung der Distanzen aus geschmolzenem Quarz nach lichtinterferometrischen Methoden, *Veröffentl. Phys. Lab. Univ. Turku*, 2, Turku, 1933.
- Väisälä, Y., Die Anwendung der Lichtinterferenzen zu Längenmessungen auf grösseren Distanzen, *Veröffentl. Finn. Geod. Inst.* 2, Helsinki, 1923.

Canadian Shoran Project

Abstract

J. E. R. ROSS

*Geodetic Survey of Canada
Ottawa, Canada*

The Canadian Trilateration Survey, initiated in 1949 and completed in 1957, covers an area of 2,500,000 square miles. The Shoran net consists of 143 stations (18 fixed) with 502 measured interlacing lines, of which the longest is 100 miles.

Through the Shoran trilateration Canada now has a complete coverage of the area north of the present triangulation system. The Arctic lands are integrated with the mainland, and all results are upon a uniform datum. The standard accuracy in length is assessed as 1 in 56,000, and this value has been confirmed by a limited number of Shoran stations since incorporated into the triangulation net. Variations in position of some magnitude are bound to be revealed, because, although length is maintained on the average to a high degree, random mismeasurement does occur and has remained undetected during the operation. The effect of mispositioning a station is to introduce a local twist or kink in the net, and this, in turn, is imposed on the succeeding work. Owing to the location of the 18 basic triangulation stations, the distortion

in position cannot be removed now, but it cannot exceed certain limits because of the confinement of the work by these stations. An improvement in positioning and general accuracy will be obtained through the projection of first-order triangulation northward, and the geodetic value of the trilateration will then be correspondingly increased.

The Shoran trilateration has served its original purpose during the period of projection and also for a further two years. The embracive control coverage now permits a coordinated breaking down of the net in local areas by newer and shorter-range electronic instruments. Shoran, replaced by Hiran, which, in turn, will perforce give way to other instrumentation, has given Canada one of the essential bases for accurate mapping.

The trilateration has provided an important link in the connection of continents and thus has aided in the comparative study of datums; perhaps it will contribute to the selection of a world spheroid and datum.

Hiran Instrumental Developments

PAUL W. JORDAN

*Aero Service Corporation
Philadelphia, Pennsylvania*

Although the Shoran-Hiran system and its application to accurate trilateration have been covered many times and are well outlined in *Principles of Shoran Mapping* by Carl Aslakson, a brief condensation seems necessary here in the interest of clarity and in forming a background for those unfamiliar with the Shoran equipment. Without it, details of electronic distance-measurement techniques would have little interest or meaning.

A basic Shoran system consists of three components: the 'airborne' or mobile set, and two fixed (but portable) ground stations. The mobile unit can also be carried by truck or ship. The ground stations are responder beacons and merely amplify and retransmit signals of RF energy which they receive from the airborne set. These stations are alternately interrogated by the A/B (airborne) equipment, which consists of a transmitter, receiver, and indicating-timing device. The time between interrogating pulse and the returned response is precisely measured and interpreted numerically as physical separation of air and ground equipment.

Thus, the function of the ground station is to receive and amplify signals directed to it by the A/B unit and retransmit them; the function of the A/B unit is to give continuous distance readings from the aircraft to a pair of ground stations and record these readings. As the plane moves along its path, distance information is presented by two rotating dials with a resolution of 0.0002 mile, or 1 ft. When the operator is 'tracking,' each dial shows the distance at any instant to its respective ground station. Thus, we have a 'continuous fix' of aircraft position. (It must be remembered that Shoran distances are slant heights, and altitude must be recorded accurately for data reduction.)

Instead of measuring time, or distance, from transmitted to received pulse along a sweep on

the X or Y axis of a CRT (Cathode-ray tube) as does most navigational equipment, the Shoran transmitter pulse is advanced ahead of zero or reference time by an amount just equal to the time required for the energy to travel from the plane to the ground station, through the ground equipment and back to the aircraft. The response, then, will arrive back at such time as to be in exact alignment with a reference or marker pulse, and this trigger 'advance' which was a function of time, will be presented as distance information. Here is the secret of Shoran's increased accuracy over other existing systems. Since neither transmitted nor received pulses are presented on a long continuous sweep, presentation can be limited to the duration of time in which the response pulse occurs. The part of the sweep can be expanded, then, to give desired resolution and accuracy consistent with the ultimate possible from the phase advance device. (Since finite time is consumed in traveling through the ground station it must be included in the measurement, and the A/B zero can be adjusted to compensate for it.)

Ground stations are interrogated alternately one on each of two different frequencies. Both reply on a third, or common, frequency. Because of the persistence of the A/B scope, each response appears to be continuous.

When both the phase advance devices (goniometers) in the A/B unit are adjusted correctly, there is but a single point of alignment of the marker and the response from each of the ground stations. Because the operator's attention is directed to a single point, simultaneous alignment can be easily accomplished. Conversely, adjustments at separated positions along an X or Y axis would be virtually impossible.

To identify the responses on the CRT, the polarity of the presentation is alternated

chronization with the changing of the inter-
ation frequencies.

n order that the dials mentioned above read
rectly at every instant as the aircraft moves
ng its course, the operator rotates the two
meters with the aid of a mechanical drive
a rate corresponding to the speed of the
ne. Controlling this shift in phase thus al-
s him to maintain alignment of the received
ses with the marker on the CRT at all
es.

as mentioned before, the magnitude of this
ase shift is presented by these dials as nu-
rical distance information. If we choose a
ing frequency so that the duration of one
le is equal to the time required for radio
rgy to travel to a point 1 mile away and
urn, 360 electrical degrees of the timing
quency represents 1 mile. It also represents
physical degrees of rotation of the phase
fters. The phase shifter position, through
per gearing to indicator dials, indicates the
ysical separation between A/B and ground
tions in multiples and submultiples of miles.
This same timing frequency is used for CRT
eep voltages, a very stable presentation is
ssible.

Aero's Hiran system is a Shoran system
ose equipment has been redesigned specific-
y for geodetic distance measurement. The
A equipment was devised so as to take
vantage of the continuous-fix feature as an
to blind bombing as well. Aero's Hiran en-
avors to take into account most of the hi-
uracy modifications to the original Shoran
ar that have been developed over the years,
d to correct deficiencies wherever possible.
When development of the present equipment
gan, a decision had to be made whether the
ventional approach or one of several sug-
sted unorthodox methods should be pursued.
'unorthodox' here means only that these meth-
s did not conform in essence to the basic
ran system as now used.

The decision was *not* to depart drastically
om the present system, for two reasons. First,
was thought that, for the equipment to gain
ady acceptance without very extensive and
stly field demonstrations, there should be no
dical departures from existing units. Second,
e areas in which improvement to the existing

type of equipment might be made had by no
means been exhausted. There is still a great
deal to be learned about the source of existing
errors and their control. The use of Hiran in
accurate long-distance geodetic measurement is,
in the opinion of this speaker, not about to be
challenged or superseded in the foreseeable fu-
ture by any other gear. The Hiran system is
unique, and the basic design more than satis-
factory. Therefore, rather than start at the be-
ginning with an entirely new concept and make
the inevitable mistakes that accompany a pe-
riod of development and trial, it was decided
to start with existing equipment and to retain
the basic features and techniques which had
been well proved, so that the benefits of past
experience (both that of Aero Service Corpora-
tion and that of the 1370th Photo-Mapping
Group of the AF) would not be lost. In this
way problem areas could be investigated and a
reasonable number of innovations incorporated
with the greatest chance of success.

Aero Service Corporation has purchased a
number of new A/B units, the APN-84, from
RCA. It has also acquired a number of Hiran
CPN-2A ground stations, though they were
not purchased until after development of a
ground station of Aero's own design was begun.
The new units were completed, however, to en-
sure a future supply of equipment by having a
satisfactory, tested design that could be dupli-
cated as the need arose or if the benefits
from design improvements warranted additional
manufacture.

The Aero Service ground station, the CPN-
X-1, uses the same basic approach as the CPN-
2A to fulfill the function of responder and
to establish and control its own inherent delay.
At the same time, new circuitry has been in-
corporated where it was considered that im-
provement could be made. The goals were (1)
to equal or surpass the accuracy of the CPN-
2A, at the same time improving the stabil-
ity, ease of adjustment, and consistency of
measurement; and (2) to reduce the bulk and
weight of the ground-station components. Re-
packaging the transmitter was thought desir-
able to facilitate back packing and donkey
packing; many sweat-filled hours of field ex-
perience had demonstrated that the CPN-2A

transmitter was far too heavy and unwieldy as a unit.

To visualize what elements need investigation and/or circuitry improvement, let us again consider the complete Hiran system as a chain each link of which contributes its share of the distance measurement. The delay of each element must be maintained at a known and unchanging value. Any part of this air-ground loop that tends to restrict or alter the pulse rise time will in essence cause an increase in the distance measurement. Also, any element that is not stable, or causes a change in the path traveled by the pulse, obviously results in an error in measurement.

A great deal has been said in the past about 'intensity' error or distance error introduced by variations in signal strength. The leading edge of all pulses has finite rise time, and it must have, no matter how much we should like it to be otherwise. If such a pulse is used to trigger a circuit which takes anything but zero voltage to actuate it, any variation in pulse rise time will cause a shift in the trigger point. This change in trigger timing can also be brought about by an over-all amplitude change such as occurs when signal strength varies.

Let us start with the transmitter as a specific component. Change in its triggering point can be reflected as a 'delay' change and thus as an over-all distance change in the system. The apparent trigger point of the transmitter is somewhere between 0.5 and 0.75 volt. This low but definite trigger level seems to occur, not because any of the early stages are biased at, or beyond, cutoff, but because any signals much smaller than this receive insufficient amplification. Their rate of voltage rise determines pulse shape and amplitude, which in turn must overcome the fixed bias in the driver and modulator stages. A plot of transmitter trigger amplitude vs. trigger timing (or delay) shows a very distinct knee in the region of 0.5 to 0.75 volt. A lesser, but additional, delay shift is observed from this point on up to a trigger value of 3 or 4 volts, where the delay asymptotically approaches a constant value. In essence, it might be said that for lowest trigger levels (up to about 2 volts) it is the rate of change of the leading edge that determines the trigger point. As the amplitude increases beyond the 2-volt

level, the integrating characteristics of the transmitter amplifiers and driving transformer begin to take over. The entire unit finally acts more as a switch than as an amplifier above 3 or 4 volts. Amplitude change of the trigger from here on has but little effect on the unit delay.

Reducing the trigger level required by the transmitter to near zero might seem at first a logical approach to eliminating delay change due to amplitude variation, but then 'no triggering' could not be avoided, and so a compromise must be reached.

'Gain riding' was introduced as an attempt to eliminate variations in delay by reducing to a constant level all triggers fed to the transmitter, regardless of whether they come from the receiver during responder operation or from the monitor during zeroing (self-triggering).

In the CPN-2A, the transmitter trigger level was found to be so low as to increase the difficulty of gain riding. This made the equipment hypersensitive and subject to what was considered to be unnecessary variation. Therefore a great deal of thought and time went into improving receiver and monitor circuitry. The object was to decrease the rise time, increase pulse amplitude at gain-riding level, and eliminate interaction and variables in the unit. Faster trigger rise times (less than one-half of the usual 0.1 microsecond) can be produced in the new monitor, and square pulses can be presented with minimum distortion on its oscilloscope. The monitor, then, can match any transmitter or receiver development visualized for the future without redesign. (It should be mentioned that a trigger initiated in the monitor itself is matched with one that has traveled through the transmitter and receiver during 'zeroing' operation.)

An interesting side feature is the adaptability of the monitor scope for use as a trouble-shooting and test instrument in the field by proper positioning of the function switches.

One of the major problems in making a correct Hiran measurement is, logically enough, obtaining of an accurately established reference or starting point for the A/B reading. It must be subject to little or no variation, and its value must be known. This 'zero' point is influenced by both the A/B and the ground un-

and the factors affecting each are similar. The ground-station delay segment of the distance measurement, for instance, must be considered part of the zero reading. (The zero reading determines the part of the measured distance that is accounted for by the equipment itself, both air and ground, and that must be removed in order to obtain the distance attributable to physical separation of the units. This can be done electronically by shifting the original timing point of the A/B set, or mathematically by a data correction, or by a combination of both.)

During the ground-station zeroing operation, the transmitter is triggered by a pulse originating in the ground monitor itself. The time taken for the pulse to travel the path through the antennas, through the receiver, and back to the transmitter is set against a crystal-controlled time base. This is the usual approach. However, it assumes that the RF energy takes the path as outlined—from transmitter, out the transmitter antenna cable, down the receiver antenna cable, and through the receiver, where it is finally detected and presented as a video pulse to the rest of the unit. It assumes, also, that, once this delay is established, it remains constant whether the unit is in the 'zero' function or the responder function. It assumes that delay is independent of physical location of components and cable arrangement that might be shifted during any normal operation. Unfortunately, however, it is very difficult, because of the magnitude of the oscillator burst, to keep RF energy out of the interconnect cable along which it travels from the transmitter to the monitor. The receiver is sensitive enough to pick up some of the radio frequency once it is inside the monitor case. Thus, it will be amplified and presented to the scope too soon. The pulse that traveled the longer path will be 'ramped' out and will not be seen. What determines the unit delay, then, will be the length of the 'short cut' and will be altered when the unit is switched from 'zero' to transponder or responder. Also, physical location of cables with respect to sources of extraneous leakage and radiation will cause delay changes.

Along with the physical changes to the ground-station transmitter designed to reduce size and weight, the oscillator unit was com-

pletely separated from the power-supply chassis and power-switching section. Careful filtering of power leads, both leading to and leaving the transmitter, has been accomplished. 'In-line' filter units, placed in a shielded section immediately adjacent to the interconnect AN connector, reduced radio frequency present on these leads outside of the transmitter by an average factor of 70 db. Better shielding inside the transmitter itself serves as additional isolation of the power oscillator.

The use of triaxial cable for the antenna leads will also tend to reduce external radiation which can be picked up by the shields and conductors of the interconnect cable.

Careful termination of all coaxial trigger and antenna leads in the characteristic impedance of the coax will help eliminate pulse reflection, which if not eliminated will cause a change in the apparent velocity of propagation along these cables. Originally, this was one of the sources of greatest variation in ground 'loop' calibration tests. (These are lab bench tests which supersede the costly, time-consuming truck runs formerly used to verify the magnitude of the ground-station delay.)

To return to a previous statement, it was said that the transmitter was definitely an amplitude-sensitive instrument, but the receiver is the component causing the wave-form changes that eventually manifest themselves as delay changes via the transmitter's characteristics. The remaining improvements, then, seem to center around the reduction of variations in the receiver output pulse. First come improvements in the inherent electrical characteristics of the receiver itself, and second, 'gain riding' to control all residual variation.

Quite a number of Hiran receivers were purchased, but they were our biggest disappointment. Their dynamic range was completely unsatisfactory, and unless signals were reduced to near gain-ride level, the intermediate frequency quickly overloaded and the stronger signals almost completely disappeared. This might be quite tolerable for line crossing work, but on Hiran-controlled photo lines, where new stations must be quickly spotted and changeovers made 'on line,' a greater dynamic range was considered essential.

Also, to ease the task and increase the accu-

curacy of gain riding, through any possible decrease in rise time, a wider bandwidth than 7 Mc/s seemed in order. This would remove the receiver as a present or future restricting link in the measurement chain.

Consequently Linear Equipment Labs (LEL) of Long Island was commissioned to design a new IF strip similar to some they were producing for use in missiles. The result is an IF strip, adaptable for use with all conventional types of Shoran and Hiran equipment, which, when used in combination with a good RF unit such as that contained in the Hiran, can produce maximum output with a signal input of 100 to 120 microvolts. It has a bandwidth of 12 Mc/s and a gain of 90 db. A crystal detector, video amplifier, and cathode follower can produce an output of as much as 14 volts into a 1350-ohm load.

A future Hiran system has been suggested using two airborne receivers to eliminate commutation and frequency switching. This would permit replacing manual gain riding by automatic gain control, with its obvious advantages. Therefore, to prevent early obsolescence of the new strips, LEL has tested and found this design to be very satisfactory when used with AGC.

Last but not least of the new developments was the fabrication of a new gain-riding scope, using the TS-100 oscilloscope as a basis for modification. This scope performs the same primary function as the APA-11 now in use by the 1370th Photo-Mapping Group, but, to increase resolution and control over the gain-riding, function by the operator, helipots and three-turn 'microdials' were incorporated. Control is smooth, and signal-level changes of 1 db are easily compensated for.

Development work on Aero's CPN-X-1 was begun in the spring of 1957. Two basic prototypes of the ground station, modifications to one of the APN-84's, fabrication of the TS-100

gain-riding scope, and complete bench and loop testing were completed by the following spring. In June 1958 preliminary field tests were completed over a ground calibration range in the pine barrens of New Jersey. The Aero prototype and two CPN-2A's (RCA's Hiran ground units) were used in the tests.

Although these tests were limited to ground operation, the results obtained lead to the following conclusions: (1) The Aero Hiran prototype performed within Hiran specifications as predicted from laboratory tests. It is considerably easier to operate than the CPN-2A and weighs at least 40 pounds less. (2) The CPN-2A's also performed within Hiran specifications, but adjustment was extremely critical; small changes of receiver gain settings resulted in considerable shifts in distance readings. Certain modifications should improve ease of adjustment. (3) The modified airborne set and the gain-riding equipment were very satisfactory during the test runs. These components would be subjected to more stringent tests during airborne trials, of course.

One of the problems in Hiran surveying has been the yearly increase in the cost of operation; it would seem that this new equipment provides the means of increasing accuracy, measurement and consistency of results while holding the line on costs as the direct result of fewer reruns, increased dependability of operation, and elimination of what will be unnecessary ground calibration runs. 'Hiran,' it is realized, has always implied 'laboratory tests' instead of 'truck runs,' but the statements above apply equally to what is thought of as 'Shoran' and to what is considered to be 'Hiran' surveying. It is hoped that Aero with its CPN-X-1 Hiran equipment, and the contribution just mentioned, has been able to advance the state of the art in the commercial field.

Evaluation of Hiran Networks

SAMUEL D. OWEN

*1370th Photo-Mapping Group
Air Photographic and Charting Service (MATS)
Turner Air Force Base
Albany, Georgia*

INTRODUCTION

The years since World War II have seen the development of distance-measuring equipment of sufficient accuracy for application to photodetic surveying. It has been of equal importance that the equipment designed has been of a practical nature, suitable for everyday field use. That the equipment is being used is well known, and with its use have come problems related to the characteristics of the systems. The measuring systems developed may vary considerably in their capabilities of accuracy and range, and in the uses intended for them, but many of their problems are similar. Among the more important are: (a) the determination of the figure of the net required to provide the results desired from a survey with the minimum of effort; (b) A method for the field analysis of the accuracy of the observed measurements which will allow positive knowledge of the adequacy of the work; (c) Analysis of the final results.

The solution of these problems lies in the mathematical examination of the networks before, during, and after completion of the survey. The characteristics of trilateration prohibit the use of established methods. Experience with Hiran surveys has resulted in the development of certain techniques which may be suitable with other survey systems.

STRENGTH OF FIGURE

The final results of any Hiran trilateration project are the latitudes and longitudes of the Hiran stations, and the magnitudes of the uncertainties in the positions established are, therefore, the most important factors in determining network accuracy. Before a survey begins the probable accuracy with which the

geographic coordinates will be determined should be known.

Trilateration network designs for Hiran projects are based on strength-of-figure computations patterned after a network adjustment. Unknowns solved for in a trilateration adjustment by the variation of coordinates method are shifts in latitude and longitude for the various points in the net. The probable errors of these position shifts may be computed from the formula

$$PE = PE_s(1/W)^{1/2}$$

where PE_s is the probable error of a single observed distance and W the computed weight of an equation corresponding to a shift in latitude or longitude.

The factor $(1/W)^{1/2}$ may be computed during the planning phase of a project, azimuths for the trilateration lines scaled from a map being used as the basis for the computation. If the probable accuracy of the observations is known from experience, the probable errors in latitude and longitude that will result from a specific network design can be computed. Computation of the expected uncertainties in position before the start of a planned project is of value:

(a) To ensure that the network is adequate to provide the desired results.

(b) As a guide to modifying the network. For example, if the uncertainty in the longitude of a specific point is shown to be excessive, the addition of one or more approximately east-west lines terminating in that point is indicated. All apparent weaknesses in the figure of the planned net will be revealed and can be overcome before adoption of the final network design.

(c) To allow the adoption of an economical

network figure. Excessive strength in the latitude or longitude of any point will indicate types of lines that can be eliminated. Although the computation does not identify the one optimum net design, repeated computations for various figures will, by comparison, aid in approaching the best design.

(d) To determine the necessity for the inclusion of azimuth control from an independent source and the location where the insertion of such control would be most beneficial. The effect of an azimuth line may be revealed by including an azimuth condition equation in the strength-of-figure computation.

The strength-of-figure computation described above provides the basis for the adoption of the final net design for all Hiran projects. The probable errors in latitude and longitude and the resultant probable errors in position are computed before the start of any Hiran survey, taking ± 0.0025 statute mile as the value for the probable error of a single observed distance. The ratio between the probable error in position and the distance from the starting positions for every point in the net is equated with an accuracy yardstick to determine the adequacy of the net design.

Hiran surveys are ordinarily extensive, covering a broad area. In scope they more closely resemble a national survey system, such as that of the United States, than a single arc of triangulation. In practice, it is attempted to duplicate the expected accuracy of the United States national survey system as estimated by Simmons with the formula

$$\text{Proportional part accuracy} = 1/20,000M^{1/8}$$

where M is distance in miles along the axis of the net. This formula is used with the results of the strength-of-figure computation to determine the allowable magnitude of the uncertainty in Hiran positions. The specific criterion used is:

$$PE_{pos}/M = 1/20,000M^{1/8}$$

FIELD ANALYSIS

Hiran surveys are performed by aerial survey teams composed of airborne and ground personnel, including computing elements. Missions are computed as soon as the film record-

ings are developed in order to indicate the direction of further operations. The inclusion of computing personnel allows comprehensive and logical computations to be performed in the field.

In Hiran, as in any other observing system, an occasional measurement contains a gross error. Field analysis is required to assure the accuracy of the Hiran measurements while the team is still in the project area and the ground station sites are still occupied. It is unfortunate that there is no simple method for accomplishing this; the characteristics of trilateration and the complexity of Hiran networks do not lend themselves to simplified methods.

With the lack of independent checks on the accuracy of observations, the determination of the internal consistency of the network itself must be the primary method of field analysis. This is accomplished by rigorous least squares adjustment. As soon as enough lines have been measured to form a suitable figure, the adjustment is performed and the results are examined for indications of excessive error. As more lines are measured, the adjustment is extended. It is the intention to break up the net in increments which will allow the ground station equipment at the various sites involved to remain in place until the observations have been tested. If the observations appear to be of the desired accuracy, the ground stations are moved to new sites and the survey continued.

A measurement is reobserved when it is believed that an insufficient portion of the error will be recovered as a correction in the adjustment process. Inaccurate observations are not particularly difficult to identify, although the method used is, of course, not foolproof. Two factors that indicate the magnitude of error are the values of the corrections and of the computed probable error of a single observation. Although Hiran surveys are planned on the basis of an expected value of ± 0.0025 mile for the probable error of a single observation, a figure of about ± 0.0020 mile is more often achieved in a field adjustment. It is assumed that, if the PE_s is in excess of about this magnitude, it is due to one or more measurements containing gross errors. That this is true is usually evident from the distribution of the v 's, and the offending line is easily identified.

line that is assigned the largest correction almost always the line that is most in error. The complexity of the Hiran net complicates analysis in some respects, but it certainly simplifies it, too, by revealing the approximate magnitude of the errors in the measurements. Measurements almost always prove the accuracy of the analysis, and the experience gained creates confidence in the ability to estimate the accuracy not only of the individual observation but also of the over-all survey. When the field work has been completed, the aerial surveyor is aware of the adjusted results and is assured that no resurvey will be necessary.

FINAL ANALYSIS

The basic method for analysis of a Hiran survey is the examination of the results of the adjustment of the over-all network. This adjustment will usually include data such as length and azimuth conditions that were not available during field operations.

Of primary interest is the comparison between the probable errors computed from the adjustment with the planning criterion. The expected value for PE_s of ± 0.0025 mile should not be exceeded. The ratios of the probable errors of positions to distance from the initial points in the survey should compare favorably in magnitude with those expected during the planning stage. For example, for a network extending 1000 miles, the probable errors of the terminating positions should not exceed 1 : 200, or ± 0.0050 mile. For a survey of 2000 miles in length the probable error of the termination should not be greater than 1 : 250, or ± 0.0080 mile. The usual results is that the computed probable errors in position are significantly smaller than those estimated during project planning.

The only independent checks on the accuracy of the Hiran observations come from length and azimuth conditions usually obtained from triangulation surveys. Triangulation nets are normally included as integral parts of Hiran surveys in order to increase the strength of the net. Although direct comparisons of Hiran-measured distances with ground-survey distances are often available, comparisons with azimuths are usually not completely realistic, for the net would have been made more com-

plex if the triangulation had not been available. The fit of the lengths and azimuths and the strain produced on the Hiran measurements by their inclusion in the net provide an indirect picture. Lack of v 's of excessive magnitude and lack of a systematic distribution of corrections in the area of fixed conditions should indicate that the triangulation and the Hiran form a consistent survey. Examples of the effect of superimposing azimuth and length conditions on a Hiran net follow.

(a) *Faeroes*. Two Hiran stations in the Faeroe Islands were identical with triangulation stations, and length and azimuth data from the triangulation were used to provide conditions for the adjustment of the Hiran survey. This part of the network was also adjusted in the field with no conditions held fixed except the positions of the initial points, thus providing a comparison with the Faeroes survey. The adjusted Hiran distance agreed with the triangulation distance to within the nearest 0.0001 mile. The azimuth computed from the adjusted Hiran positions differed by approximately 2 seconds from the azimuth computed from the triangulation, modified to correspond to the difference in longitude between the two datums. For a 44-mile line, a 2-second difference in azimuth corresponds to about a 2-ft difference in the relative positions of the two points. The inclusion of length and azimuth conditions for this line apparently made no significant change in the network. Its greatest value was as a check on the accuracy of the Hiran survey and the assurance it gave of accuracy.

(b) *Crete-North Africa*. Table 1 gives a comparison between ground-survey distances, observed Hiran distances, and distances resulting from an adjustment in which only a single position and azimuth were held fixed. The observed Hiran distances do not, of course, agree exactly with the ground-survey distances, but the comparison is considered reasonably close. The greatest indication of error is for line 2-3, where the difference between the two distances is 0.0033 mile. That the Hiran distances contained a certain amount of error is corroborated by the results of the free adjustment of the net, as the adjusted Hiran distances for these lines show a more favorable comparison with the ground-survey distances.

TABLE 1. Project 51 AS-25; Crete-North Africa Tie

Line No.	Ground-Survey Distance	Observed Hiran Distance	Observed Hiran-Ground Survey	Adjusted Hiran Distance, free adjustment	Adjusted Hiran-Ground Survey
1-2	95.1550	95.1568	+0.0018	95.1563	+0.0013
1-3	220.4951	220.4981	+0.0030	220.4969	+0.0018
2-3	130.0571	130.0538	-0.0033	130.0559	-0.0012
5-6	82.9631	82.9662	+0.0031	82.9649	+0.0018

The differences were reduced in every case, line 2-3, which seemed to be the most in error, changing the most in the adjustment. There is, of course, some error in the triangulation distances which would contribute to the magnitude of the differences. As the lines were relatively short, it is reasonable to assume that the inclusion of the triangulation data as fixed information tended to improve the accuracy of the adjusted lengths for the other lines in the net and that the errors in these lines are probably smaller than the differences listed above.

The performance of surveys connecting geodetic datums and the use of data from triangulation surveys referenced to local datums as conditions in network adjustments introduces the problem of the effect of datum differences on these conditions. The various datums involved will, of course, have different orientations, and the data extracted from the various surveys must be consistent with the orientation of the datum to which the Hiran survey is referenced. It should usually be sufficient merely to modify the triangulation survey azimuth to correspond to the difference between the longitudes of the local datum position and the adjusted position of the Hiran station. Lack of indicated systematic corrections to the Hiran-measured distances should indicate the accuracy of this procedure and the consistency of the conditions obtained from different datums. A more precise check on the accuracy of the procedure in any specific case may be obtained by conversion of the local datum positions to the datum of reference. The differences in latitude and longitude of the position of the initial point of a local datum and the Hiran-determined position may be used with standard formulas to convert the position of the Hiran

sites to the desired datum. The comparison between these positions and the positions resulting from the adjustment of the Hiran network would reflect the accuracy of the method employed to utilize the conditions.

The positions included in the various datum may, of course, be placed approximately on the same datum before the network solution by utilizing the preliminary positions in the conversion to a common datum. The preliminary positions used in a final adjustment of the Hiran net, which are usually obtained from field adjustments of the survey, would closely approximate the final values, and the data used for length and azimuth conditions would have been obtained from similarly oriented surveys.

APPLICATIONS TO OTHER SYSTEMS

The procedures discussed previously may be generally suitable in any distance-measuring system involving network adjustments. Modifications must, of course, be made to satisfy the requirements of various systems.

The strength-of-figure computation could, perhaps, be useful in determining the adequacy of net design and should be used with an accuracy yardstick which corresponds to the capability of the system and the accuracy of results desired. The procedure could form a basis for the development of classifications and standards of accuracy for trilateration networks in general and for making the standards of trilateration consistent with those of triangulation.

The field analysis of a Hiran network through adjustment is made relatively easy because of the complexity of the net. A strength of figure which would correspond to an economical

gn for some other system could be much complex and might not allow real insight the existence of inaccurate observations. example, it may be intended to perform ilateration survey with a distance-measuring em whose measuring accuracy would be valent to the accuracy of the length of a of a triangulation net. To duplicate the uracy of a triangulation survey it would only necessary to adopt the figure of a triangulation net, but, since there is only one redundant in a trilateration quadrilateral, the ability recognize a bad measurement would be ted. The net would simply not be of sufficient complexity to indicate the existence or gnitude of excessive error or to identify it. lack of network complexity would, how-

ever, lend itself to the use of other methods of investigation. The various triangles could be solved and the triangle closures held to tolerances similar to those specified for triangulation.

The probable errors of positions computed from a network adjustment or approximated in a strength-of-figure computation may be used in determining an equitable distribution of discrepancies between networks of trilateration and/or triangulation. A survey loop closed partly by a trilateration net and partly by a triangulation survey will have different positions at the common points if no simultaneous solution is made. The position discrepancies may be distributed according to the probable errors of the positions established by the two surveys.

Simultaneous Adjustment of Triangulation and Trilateration An Investigation of Tellurometer Lengths

L. A. GALE

Geodetic Survey of Canada, Ottawa

The present development of distance-measuring instruments suitable for geodetic surveys raises the problem of simultaneously adjusting observed lengths and angles. The basic mathematical tools for carrying out such adjustments according to the method of least squares have existed for many years. In recent years excellent material, written in terms of matrix algebra, has been published on the general theory of least squares adjustment by such authors as Marchant, Bjerhammar, Wolf, and Gotthardt. In some of this literature the approach has been to proceed from the general problem to particular cases as opposed to the synthesis of particular cases into a general method. For example, the condition equation method and the indirect observation equation method are treated as special cases of the general method.

The type of equations and the method of solving normal equations used for any adjustment depend on (1) the nature of the problem; (2) the computational facilities available; (3) the personal preference of the computer.

If we have a surplus of observations, and if we accept the method of least squares, our main problem is not the development of mathematical formulas or tools for the adjustment of data; it is the old one of assigning appropriate weights to the observed quantities.

In an effort to contribute toward a satisfactory answer to this problem as it concerns observations of directions and measurement of lengths with the Tellurometer, a field party of the Geodetic Survey of Canada measured thirty-five lines of an east-west arc of first-order triangulation extending 137 km westward from Kenora, Ontario, to Whitemouth, Manitoba. The lengths range between 13 and 43 km with a mean of 22 km. Two lines, 9-11 and 25-26, were measured with the Geodimeter (see Fig. 1). From this operation we also hoped to gain information

about (1) the reliability of the Tellurometer when operated under Canadian field conditions over a period of 3 or 4 months with no recourse to elaborate test equipment; (2) observation procedures suitable for various requirements; (3) methods of application in conjunction with theodolite measurements.

It was not economically feasible to measure all the lengths in the triangulation, but the original aim of measuring at least all the sides of a chain of triangles was achieved (see Fig. 2).

Field tests of observations included one for the consistency of the sets of measurements of each line and one in which the measured lengths were checked against the observed angles by means of the sine law. For the latter test we assumed a standard of $\sigma_a = \pm 0.85''$ for the angles and a standard relative error $\sigma_s/s = \pm (1/200,000)$ for the observed lengths. The discrepancies ΔF obtained by using observed values in the relation

$$\Delta F = a \sin B - b \sin A$$

where a and b are two sides of a triangle and A and B are the respective opposite angles, were compared with

$$\sigma_F = \pm a \sin B$$

$$\cdot [(2\sigma_s^2/s^2) (\cot^2 A + \cot^2 B) \sigma_a^2]$$

The expression for σ_F is derived from the sine law using the law of propagation of error. With the values of σ_a and σ_s/s indicated above 78 per cent of 67 tests fell within the limits given by σ_F .

As part of the investigation, the following five adjustments, in which condition equations were used, were carried out: (1) corrections applied to observed directions only (Fig. 1); (2) corrections applied to all observed directions and lengths (Fig. 1); (3) corrections applied

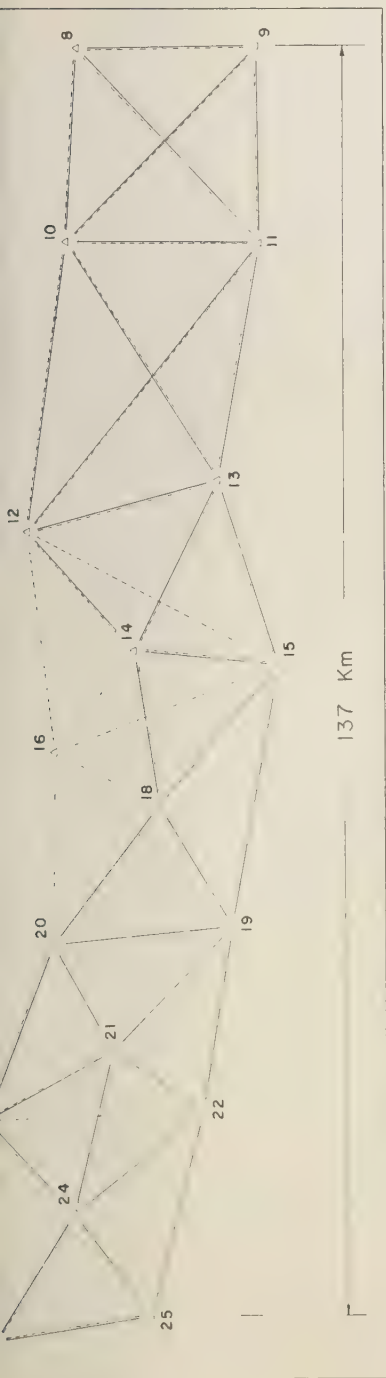


Fig. 1. Triangulation scheme. Solid line, measured lengths, dashed line, measured directions.

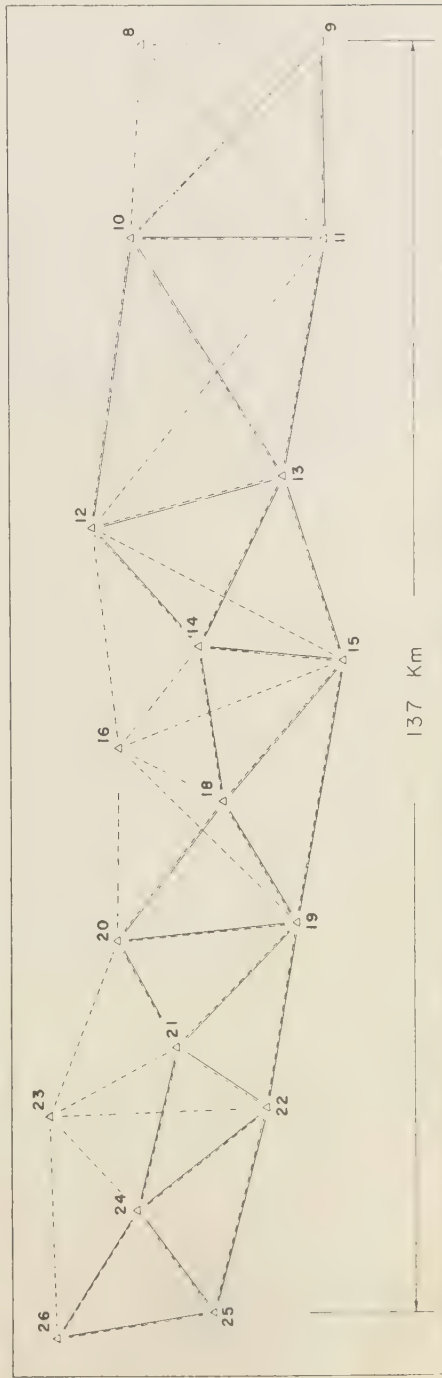


Fig. 2. General triangle scheme. Single chain used for adjustments 3, 4, and 5 shown as a solid line.

TABLE 1. Some Data and Results of Five Adjustments

Adjustment	1	2	3	4	5
No. of stations to be fixed	16	16	13	13	13
No. of conditions	37	70	39	19	20
No. of directions observed	86	86	54	54	28
No. of lengths observed		33	25	5	25
Standard error (weight unity), seconds	0.516	0.519	0.531	0.511	0.532
Average relative correction to lengths, ppm		2.9	3.5	5.4	2.6
Maximum relative correction to lengths, ppm		12.4	13.0	8.4	12.0
No. of relative corrections to lengths greater than 10 ppm		1	2	0	1
Latitude, terminal station	49°56'55.124"	55.127"	55.122"	55.125"	55.118
Longitude, terminal station	95°58'52.442"	52.472"	52.475"	52.461"	52.478
Terminal azimuth	352°21'38.992"	39.104"	39.267"	39.685"	39.079

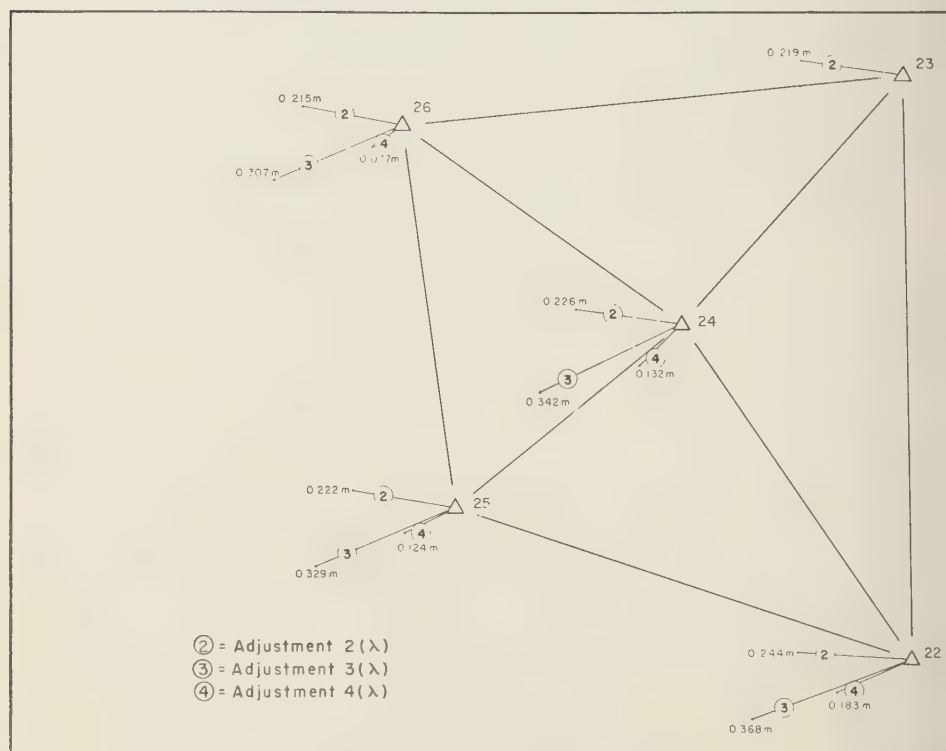


Fig. 3. Positions of terminal stations relative to adjustment 1.

to all observed directions and lengths of a chain of triangles (Fig. 2); (4) corrections applied to the observed directions of the chain of triangles and to five observed lengths along the south

side of the chain (Fig. 2); (5) corrections applied to all the observed sides of the chain of triangles and the directions from stations along the south side of the chain (Fig. 2).

TABLE 2. Lambda Values

Adjustment	λ , meters
2(λ)	-0.070
3(λ)	-0.068
4(λ)	-0.072

In all the adjustments, two Geodimeter lines, one at each end of the arc, were treated as fixed base lines.

In the adjustments, corrections were applied to the logarithms of the lengths, the units being $ds/s \times 10^6$, with $M = \log_{10} e$.

Since we assumed a constant relative error in the measured lengths, the application of logarithmic corrections allowed us to use the same weight for all the logarithmic distances. We understand that the U. S. Coast and Geodetic Survey has used $ds/(s \sin 1'')$ as the unit of correction, which for weighting purposes has the same property as the logarithmic correction when a constant relative error is assumed.

On the basis of field tests and some small preliminary adjustments, we used a weight ratio of 20 to 1 for directions with respect to logarithmic distances. Assuming $\pm 0.5''$ for the standard error of a direction, this weighting implies a standard relative error in length of approximately $\pm (1/194,000)$.

A perusal of Table 1 shows that the standard errors (weight unity) are approximately equal with a value of $\pm 0.5''$. In these adjustments the standard error represents the standard error of a direction. If we recall that no Tellurometer lengths were used in adjustment 1, we

see that the inclusion of such lengths in the subsequent adjustments did not appreciably alter the standard error of a direction.

A comparison of the adjusted lengths (not given here) shows that all the final lengths resulting from adjustment 2 are greater than the corresponding lengths of adjustment 1, one length of 3 is less than the corresponding lengths of adjustment 1, and only three lengths of adjustments 4 and 5 are shorter than the corresponding lengths of adjustment 1. This systematic discrepancy is also reflected in the longitudes of the terminal station (see Table 1), when we remember that we are dealing with an east-west arc of which the east end is fixed in position (see also Fig. 3). We also notice the tendency toward increase in the length of the triangulation scheme as the ratio of the observed lengths with respect to observed directions increases.

On the assumption that a constant systematic error had occurred in all the length measurements, we incorporated a constant correction λ for each length in the equations of adjustments 2, 3, and 4; i.e., each observed length s_i was subjected to a correction ds_i and a correction λ which was common to all observed lengths in any one adjustment. The 'lambda' adjustments corresponding to adjustments 2, 3, and 4 are designated 2(λ), 3(λ), and 4(λ). The values obtained for λ are shown in Table 2.

With reference to Table 3, which shows some of the data and results of the lambda adjustments, we note that the values of the standard error of an observation of weight unity are all approximately $\pm 0.5''$ and are in close agreement

TABLE 3. Data on Lambda Adjustments

	2(λ)	3(λ)	4(λ)
No. of stations to be fixed	16	13	13
No. of conditions	70	39	19
No. of directions observed	86	54	54
No. of lengths observed	33	25	5
Standard error (weight unity), seconds	± 0.508	± 0.519	± 0.505
Average relative correction to lengths, ppm	3.3	3.4	5.3
Maximum relative correction to length, ppm	11.5	12.0	9.2
No. of relative corrections greater than 10 ppm	1	2	0
meters	-0.070	-0.068	-0.070
Latitude, terminal station	49°56'55.126"	55.120"	55.122"
Longitude, terminal station	95°58'52.453"	52.457"	52.447"
Terminal azimuth	352°21'39.104"	39.226"	39.539"

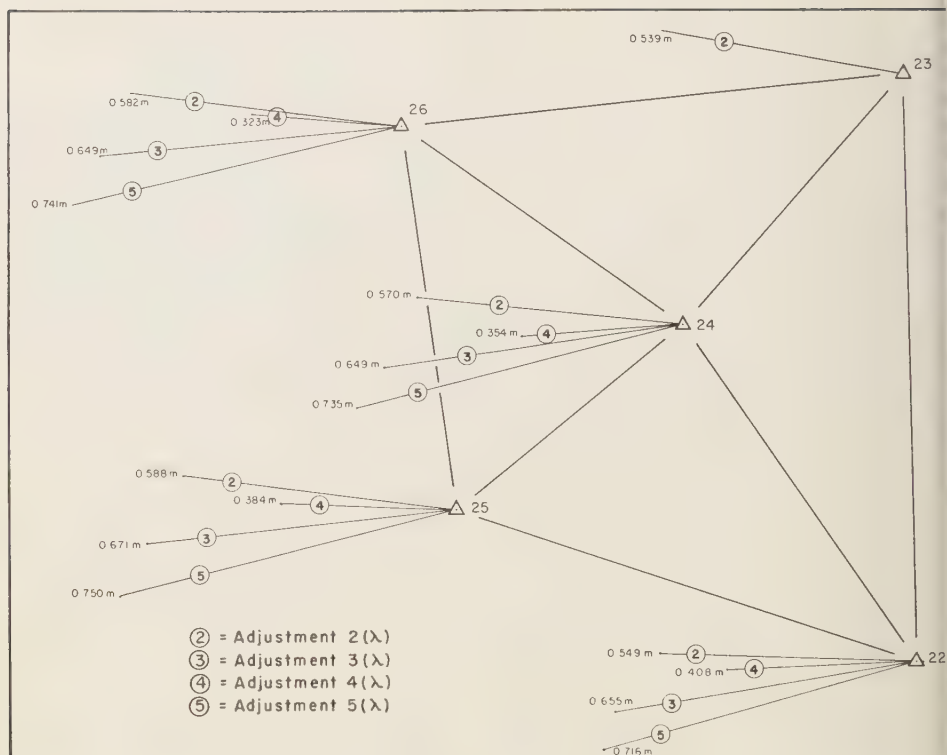


Fig. 4. Positions of terminal stations relative to adjustment 1.

with those of the previous adjustments shown in Table 1. There are also no significant discrepancies between the average relative corrections of the first adjustments 2, 3, and 4 and the corresponding lambda adjustments. The shorter lengths of the lambda adjustments are indicated by the longitudes for the terminal station (no. 25) shown in Table 3. A comparison of Figures 3 and 4 shows the effect of the lambda adjustments on the positions of stations in the terminal figure of triangulation. Figure 4 also indicates that the tendency toward greater lengths when observed lengths were included in the adjustments was not completely eliminated by the lambda adjustments.

The results of these adjustments show a definite trend toward greater lengths in the particular triangulation scheme as the ratio of the observed lengths to the observed directions is increased. For example, in adjustment 2 the

ratio of lengths to directions is 33 : 86 and the linear discrepancy in position of terminal station 25 is 0.588 meter with respect to adjustment 1, and the corresponding values for adjustments 4 and 5 are 5 : 54, 0.384 meter, and 25 : 28, 0.750 meter. This apparent overmeasurement could be a function of a number of factors, including the instrument used, the terrain, and meteorological corrections. It might also be ascribed to the direction observations, but, because of the length control at each end of the arc, the angle measurements probably contribute very little to this tendency. No conclusions can be drawn from a comparison of Tellurometer lengths and Geodimeter lengths for the two lines 9-11, and 25-26. For line 9-11 the Tellurometer length was shorter than the Geodimeter length by 0.9 ppm, and for line 25-26 the Tellurometer measurement was longer than the Geodimeter measurement by 1.4 ppm.

comparison of the standard errors of an observation of weight unity indicates, we believe, that the weight ratio of directions with respect to lengths used in these adjustments is reasonably satisfactory.

With our present lack of quantitative knowledge about factors that may significantly influence the accuracy of Tellurometer measure-

ments, we cannot rely on this instrument to provide basic length control for first-order triangulation schemes. It is our hope that more experience and more data obtained from operations in different areas by geodetic organizations may enable us to assess more accurately the factors affecting measurements with the Tellurometer.

Experimental Results of Field Tests of the Application of the Lorac Phase Comparison Radio Location System to Distance Measurement

B. W. KOEPEL

*Seiscor Manufacturing Company
Tulsa, Oklahoma*

INTRODUCTION AND PURPOSE OF PROJECT

The experimental results reported in this paper were obtained under the auspices of a U. S. Army Engineer Research and Development Laboratories project during the period February to May 1956. The objective of the project was to investigate the possible application of a phase comparison radio system to the accurate measurement of distance. As a preliminary investigation relative to this objective indicated good prospects for success in an actual field test (completed in December 1955), the project field test was scheduled for early 1956. The Louisiana Gulf Coast was selected as a test area.

The project plan of tests specified types of lines to be measured as follows. Three lines were to extend entirely over water between existing offshore oil-well platforms. One of the offshore lines was to be not less than 8 miles from land at any point and at least 100 statute miles in length. Three lines were to be entirely over land of various terrain conditions to determine their effects on the accuracy of the measurement. A variety of lines was to be measured over various combinations of land and water to provide information about their effects. A number of lines of various lengths were to be measured extending to such distances that the practical range could be determined in relation to the specified measuring radio frequency. The program of lines thus proposed by the plan of tests is shown in Figure 1.

FUNDAMENTAL METHOD OF DISTANCE MEASUREMENT

In a phase comparison position-determining system the geographic positions of the system transmitters are determined to a high degree of

accuracy from existing precise survey information. From this information the earth surface base-line orientation and length between a pair of transmitters in the positioning system may be computed. The base-line length is then divided into N intervals, each equal to one-half wavelength of the positioning system's base radio frequency. The half-wavelength unit is called a *lane* and is 1 unit or revolution of phase comparison measuring instrument. In this way the number of lanes N in a base line becomes known. In the distance-measuring problem the geographical positions of the transmitters on the ends of the base line are known only approximately from astronomic observations for position or by scaling from existing maps. The width of a lane can be known to accuracy which is dependent on the accuracy to which the basic radio frequency of the base line transmitters and the velocity of propagation of electromagnetic radiation in the earth's atmosphere are known. The number of lanes in the unknown base line is determined by the ratio of the true (but unknown) base-line length to the known lane width. The base-line length may then be determined if the number of lanes in the base line can be counted.

The lanes can be counted by circulating a phase comparison lane-counting receiver from one base-line extension to the opposite end. The actual path traced by the receiver in passing from one base-line extension to the opposite one is entirely arbitrary as long as the receiver stays in range of the base-line transmitters. The word picture drawn above is shown graphically in Figure 2. In this figure the base-line transmitters are labeled T_1 and T_2 . The straight line between these transmitters is the base line of unknown length. The curved lines which converge to cross the base line are hyperbolic lanes of constant phase value or lane contours.

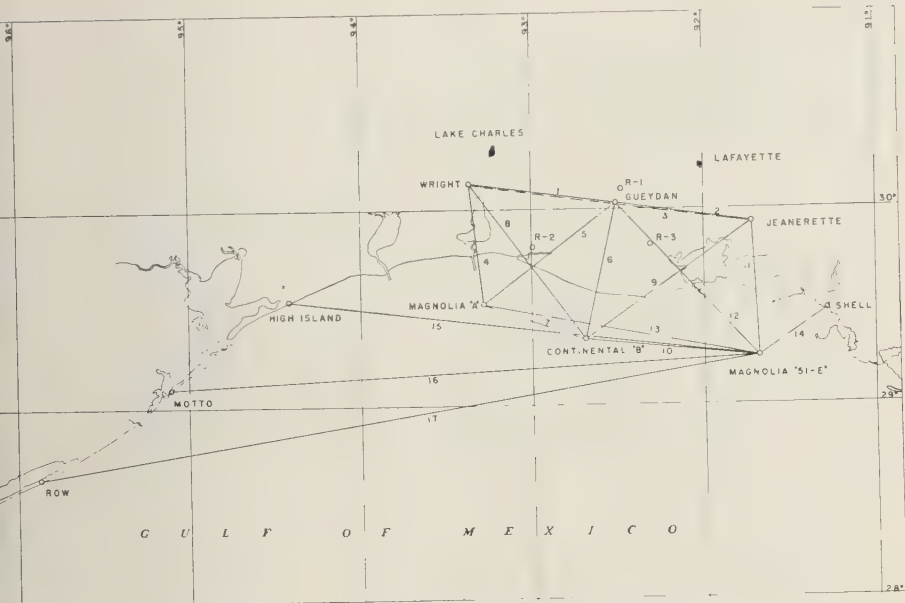


Fig. 1. Test area, February-May, 1956.

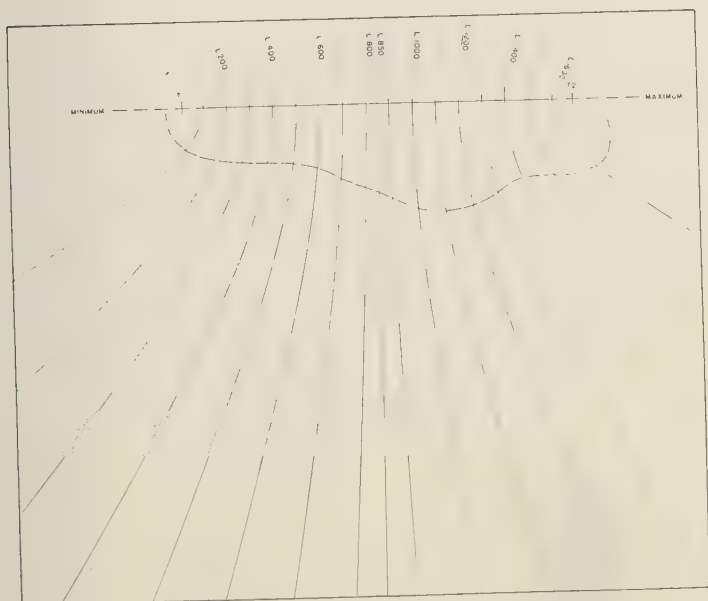


Fig. 2. The principle of lane counting along an arbitrary path.

The dotted line from the base-line extension extending to the left of station T_1 to the base-line extension extending to the right of station T_2 is seen to cross each lane regardless of the route selected to pass from the one point to the other. A lane-counting receiver following the path shown would provide an integrated lane count plus fraction for the base-line length L .

For a particular selection of radio frequencies, the phase comparison receiver will indicate a minimum reading on the base-line extension to the left of T_1 and a maximum reading on the extension to the right of T_2 . The difference in lanes between the minimum and maximum phase comparison readings is just equal to the number of lanes in the unknown base line. The fact that phase comparison observations are actually made at only two points in the system has led to the development of methods of line measuring where it is not necessary actually to trace out the long path from one extension to the other. This method employs multiple measuring frequencies and will be described in greater detail later.

The unknown line length may be expressed, in simplified form, in terms of the variables upon which it depends, as follows:

$$L = \frac{2\mu(a)f[\psi_{\max} - \psi_{\min}]}{C} \quad (1)$$

where ψ_{\max} and ψ_{\min} are the phase comparison receiver readings on the respective base-line extensions. The refractive index of the earth's atmosphere at the surface of the earth is de-

noted by $\mu(a)$. The measuring radio frequency is f , and the velocity of propagation of electromagnetic radiation in free space is represented by C .

As one might suspect, a number of natural factors are actually operating at all times which tend to disturb the simplicity of equation 1 by the introduction of more complex variables, particularly when precise results are expected for L . These factors will be divided into (1) instrumental, (2) propagational, and (3) geometrical factors. They are the subject of the next section.

FACTORS INVOLVED IN THE ACCURACY OF PHASE COMPARISON MEASUREMENTS

In considering the various groups of factors affecting the accuracy of a line length measurement, it will be necessary to return from time to time to the basic operation of a phase comparison system itself. To facilitate these investigations a block diagram of a phase comparison line measuring system is shown in Figure 3. Its operation may be briefly described as follows. Radio-frequency carriers f_1 and f_2 , which are different by only a very small frequency, are radiated from transmitters T_1 and T_2 . Since these radiations are nearly equal in frequency they are readily received by the receiver R_1 in the phase comparison measuring unit. As a result of a process known as 'heterodyning,' or 'beating,' the radio-frequency carriers within the receiver, the signal which emerges from R_1 is actually the difference frequency ($f_1 - f_2$). It is usually more conveniently represented by the letter n_p .

The frequency n_p is arbitrarily chosen, usually in the range from 100 to 600 cycles per second. Its phase is dependent directly on the difference in phases of the two radiated carriers f_1 and f_2 . The phases of these two carriers are to a high order of approximation, linearly related to the transmission distances r_1 and r_2 . Figure 3. In this way it can be seen that the phase of the signal n_p is directly proportional to the difference in distance ($r_2 - r_1$). Before the phase of the difference signal n_p can be measured, it is necessary to provide a reference signal of the same frequency as n_p and constant in phase. With the reference signal at hand it is possible to compare the position-sensitive

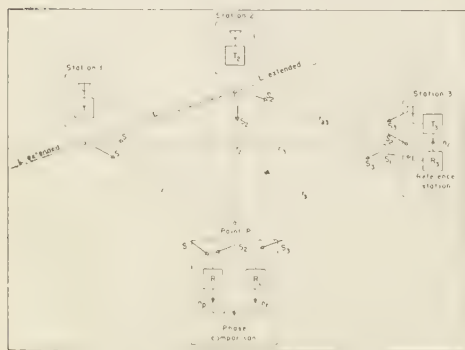


Fig. 3. Diagram of phase comparison line measuring system.

al n_r , with the reference signal and display result of the comparison.

Referring to Figure 3, it can be seen that a reference frequency signal n_r will be developed at the output of receiver R_3 , which is identical to the output of receiver R_1 , and that the phase of this signal will be constant since the transmission distances to the reference station receiver R_3 are all constant. The constant phase reference signal n_r is impressed on the carrier of the transmitter T_3 and radiated to the receiver R_2 of the phase comparison measuring unit. Although it is not feasible to radiate the signal directly, its conduction to receiver R_2 by means of the carrier f_s is for practical purposes equivalent to the direct radiation of n_r itself. Now, the frequencies f_1 and f_s are usually of the order of magnitude of several millions of cycles per second, whereas n is of the order of hundreds of cycles per second; consequently the difference in wavelength between f_1 and n is very large. The transmission of the signal from reference station 3 thus provides a reference signal at the output of receiver R_2 which is virtually insensitive to variation in position of the phase comparison measuring unit.

The phase sensitivity of the position signal can be estimated from the following considerations. If the measuring frequency f_1 is 2 Mc/s, 1 wavelength or 360 electrical degrees of phase of this carrier is 492 ft. For $n = 100$ cycles per second 1 wavelength is 9,835,705 ft. Small changes in position of the phase comparison measuring unit produce approximately equal and opposite phase changes in the two carriers from stations T_1 and T_2 , the difference phase change is twice the phase change produced in one carrier. Thus, when the position of the measuring unit changes by an amount equal to one-half wavelength of the carrier f_1 , the phase of the signal n_r changes 1 full revolution, or 360 electrical degrees. Thus a lane is 3 ft wide on the base line using a measuring frequency of 2 Mc/s. The phase measuring scale is usually graduated into 100 units so that each unit represents 2.46 ft on the base line. The special line measuring phase comparison receiver was modified to read to the nearest one-thousandth of a lane, or 0.246 ft. This amounts to about 3 in., or 250 micromicrosec-

onds if converted to a direct time measurement.

With the above explanation of the operation of the phase comparison radio system, we proceed with the examination of factors that must be considered in precise line length measurements.

Instrumentation

In equation 1 above, the measuring frequency f_1 occurs, and the line length L depends directly on this factor. During the project tests the frequency f_1 was developed by a quartz crystal oscillator of routine design. Oscillators of this type may be depended upon to hold frequency within ± 1 cycle per megacycle per degree centigrade of ambient temperature. In order to reduce the variation in measuring frequency to the minimum amount, the project crystal oscillators were operated in thermostatically controlled ovens so that temperature was maintained to within $\pm 1^\circ\text{C}$. In this case a measuring frequency of 2 Mc/s, a 2-cycle variation from the absolute frequency, amounts to a variation of $\pm 0.4918 \times 10^{-3}$ ft per lane. In a line length enclosing 2000 lanes, or about 100 miles long, the variation in measuring frequency would produce a variation of about ± 1 ft in the measured line length. Further improvement in line length measurement results was achieved by comparing the project measuring frequency continuously against one of the primary standard frequencies emitted by radio station WWV at Arlington, Virginia. The exact mechanics of this measurement is not difficult. The result of the measurement was the absolute value of the measuring frequency to ± 0.1 cycle per second. Variations of this order produce variations of $\pm 0.2459 \times 10^{-4}$ ft per lane, or about ± 0.05 ft in 2000 lanes. As a result of the control exercised over the measuring frequency, errors due to this source were considered to be reduced to an insignificant value.

Beat frequency control. Although the beat frequency n or $(f_1 - f_2)$ does not appear in equation 1, it does appear in some low-ordered terms of the fully expanded form of that equation, which was used for actual line length computations. Furthermore, if n is allowed to vary, differential phase variations at the final phase comparison point may appear, owing to the dissimilar paths followed by signals from

the base-line transmitters to the outputs of receivers R_1 and R_2 . For these reasons the beat frequency was monitored at station T_2 and compared with a standard signal of frequency n . The standard signal was developed by a vibrating reed capable of maintaining frequency to within 1 part in 20,000. With the beat frequency controlled to this accuracy, differential phase shifts in the instrumentation become negligible. Since the sensitivity of the line length measurement to variations in n is extremely small compared with the sensitivity of the line length result of f_1 , errors contributed by the beat frequency n will be insignificant.

Station T_3 phase control. The two halves of the receiving system at the phase comparison point are identical and consequently have phase characteristics, relative to most variables, which produce very small differential phase errors. The reference signal n , however, must be developed in receiver R_3 at station T_3 , go through the process of modulation onto the carrier of transmitter T_3 , and, finally, radiate to receiver R_2 at the phase comparison point. Now, although the control of the beat frequency n to narrow frequency tolerances does much to eliminate phase errors due to variation of the frequency n in the circuitry of this station, some phase shifts may occur for other reasons. To reduce these errors to the minimum, a sample of the signal radiated from the transmitting antenna at T_3 was demodulated and phase-compared with the signal out of receiver R_3 . Any phase variations produced by the circuitry in station T_3 for reasons other than frequency variations were thus detected and eliminated by automatic equipment or by the station operator.

Propagation of Electromagnetic Energy

The propagation of the signals emitted by transmitters T_1 and T_2 has been assumed to be a linear function of distance in considerations up to this point and in equation 1. Actually the RF carriers experience effects produced by the earth's atmosphere through which they must pass and the earth's surface materials over which they must pass. Because of the tedious mathematical details that must be considered to fully appreciate the effects produced in a radio carrier being propagated over the

earth's surface, the various effects will be mentioned and described only.

Free-space velocity of EM propagation. The limiting velocity of electromagnetic radiation is its velocity in free space. This velocity has been the subject of a great many investigations in the past and continues to be a matter of concern. During the past ten years all the determinations of free-space velocity have been in good agreement, and a number of them have been combined with a group of other, measured and interrelated, physical constants in a least squares determination of their best values. The value thus computed for the free-space velocity of electromagnetic radiation is

$$C = 299,792.9 \pm 1.6 \text{ km/sec}$$

This value for C is used throughout all computations.

Refractive index of the earth's atmosphere. The effective wavelength, or lane width, of the phase comparison radio signals in the earth's atmosphere depends on the free-space velocity of propagation of electromagnetic radiation, frequency, and the index of refraction of air. A value for the free-space velocity of propagation was considered in the preceding section. The measurement of the signal frequency was considered above under 'Instrumentation.'

The index of refraction of any material may be defined as the ratio of the velocity of propagation of electromagnetic radiation in free space to the velocity of propagation in the material. The properties of air on which the index of refraction depend are its pressure, temperature, and water-vapor content. The equation relating these quantities and the index of refraction is

$$N = (n_{t,p} - 1) \times 10^6 \\ = (77.6/T)p + 4810(e/T) \quad (2)$$

where

- p = total air pressure, mb.
- e = partial water-vapor pressure, mb.
- t = temperature, °C.
- T = temperature, °A.
- N = scaled-up refractive index.
- n = numerical value of index.

The development of this form is given in R

arch Paper 2385 of the National Bureau of Standards.

Air pressure, temperature, and humidity were measured continuously during the project field tests at five locations, three of which were T_1 , T_2 , and T_3 shown in Figure 3 and the remaining two were on the base-line extensions at the points where crossings were made by the phase comparison receiver unit. From the measured atmospheric or meteorological data, average values of the index of refraction can be computed for any time along all the pertinent paths in Figure 3.

The inclusion of the index of refraction in computational forms so far has been only in the linear relationship of a radio carrier phase with distance or the effective wavelength. Instantaneous values of refractive index are used in the determination of effective wavelength, as the linear phase angle of a carrier is most sensitive to small changes in the refractive index of the atmosphere. The index of refraction is also required in the equations of *nonlinear phase delay* with distance. These equations also take into account the electrical characteristics of the earth's surface. As the equations for nonlinear phase delay are relatively insensitive to changes in the index of refraction a mean index of refraction is used at this point.

Since the air pressure, temperature, and water vapor content all vary with altitude, the index of refraction of the atmosphere is a function of altitude. Whereas the index of refraction profile with altitude varies considerably from hour to hour and day to day, the index of refraction generally decreases with increasing altitude.

The radiation emitted from the transmitting antennas in the phase comparison system may be represented as a series of concentric spherical wave fronts. Lines perpendicular to these wave fronts and passing through the origin of radiation are called rays. Since the refractive index of a moist atmosphere decreases with altitude, the upper parts of the wave-front surface move with a greater velocity than the nearly horizontal parts. This has the over-all effect of causing the rays to be bent downward so that radiation emitted at low angles of elevation eventually returns to the earth's surface at some distance from the transmitter. The curva-

ture of the rays by the atmosphere is called vertical refraction.

In an average atmosphere and in the lowest few kilometers where most short-wave propagation takes place, it may be assumed that the decrease of refractive index with height is linear, though the rate of decrease depends on weather conditions. It can be shown by further analysis that a linear lapse rate of refractive index with altitude may be absorbed in calculations by using a modified value for the radius of the earth. By this means it is possible to eliminate the refraction effects produced by vertical refraction, and the earth's atmosphere may be treated as if it were homogeneous.

The modified earth radius will be referred to again in connection with geometric corrections to line measurement data. The modified earth radius factor also appears in the nonlinear phase delay equations, but a mean value for it is used there.

Nonlinear carrier phase delays. Owing to interaction between electromagnetic radiation in space and conducting boundaries such as the earth's surface, adjustment is constantly taking place in the phase of the radiation as it progresses along the boundary face. As a result, the phase of an individual carrier is not a linear function of distance, as was assumed in equation 1.

The interaction effects mentioned above are manifest in the carrier by causing it to require more time in transmission between two points than the transit time given by r/V . The actual delay in time or equivalent lag in phase angle is a function of (1) the vertical lapse rate of refractive index of the earth's atmosphere, (2) the earth's surface value of atmospheric refractive index, (3) the electrical conductivity and dielectric constant of the earth's surface, (4) the frequency of the radiated energy, and (5) the distance over which radiation takes place.

Since the development of applicable mathematical forms is well covered in the literature such information will not be repeated here. Generally the nonlinear function of phase-lag versus distance is exponential in nature and becomes linear with distance for large distances.

A problem that arises in the application of phase comparison radio systems to distance

measurement is that the conductivity and dielectric constant of the earth's surface vary radically at times, from one region to another. Several solutions to this problem have been advanced by different investigators. Where a multiplicity of different regions occurs in the signal paths, all with different or alternating electrical constants, reduced accuracy can be expected in the nonlinear phase lag corrections computed for such a path. The accuracy of the correction remains good, however, for most mixed path conditions.

Geometrical Corrections

Inasmuch as all transmitting and receiving antennas in the phase comparison radio system are elevated above the surface of the earth, the measured line must be transferred from the phase centers of these antennas to the earth's surface. Since corrections of this type depend only on the geometric shape of the earth and the antenna system, they are considered separately as geometric corrections.

Altitude correction. No mention has been made so far of the means used to transport the phase comparison receiver from one base-line extension to the other. In six separate surveys made to determine various base-line lengths, three have been conducted with surface ships and three with aircraft. In passing across the base-line extension it usually happens that the phase comparison receiver antenna does not exactly intersect the base-line extension but passes either slightly above or possibly below it. The nonintersection of the base-line extension is equivalent to not counting all the lanes in the base line or neglecting small fractions of

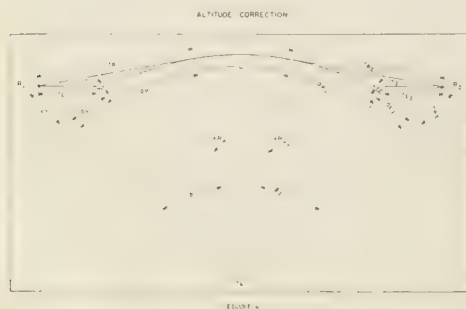


Fig. 4. Altitude correction.



Fig. 5. Sea-level correction.

the lane on each end of a base line. The correction arising from this source is very small and is insensitive to the factors involved in the computation. Since the correction arises from the altitude of the receiving antenna above the earth's surface it is called the 'altitude correction.' It becomes sensitive with increasing altitude and short distances from the extension crossing point to the nearest phase comparison transmitter. For these reasons, extension crossings were made approximately 10 miles from the nearest transmitting station and at an altitude of 500 ft. The geometry involved in the computation of this correction is shown in Figure 4. Straight-line ray paths between transmitting and receiving antennas are based on the modified earth radius factor k , which appears in Figure 4 multiplied by the earth radius. This factor was discussed previously.

Sea-level correction. When all corrections described so far are applied to the line measurement data the line length turns out to be the distance from the phase center of the transmitting antenna to the phase center of the opposite one. To reduce the measured line length to sea level a correction is determined which is the difference between the transmitting antenna phase center to horizon distance and the distance along the surface of the earth to the horizon. The geometry involved in this computation is shown in Figure 5.

DESCRIPTION OF FIELD METHODS AND RESULTS

The project test line program is shown in Figure 1. The phase comparison receiving system used to make precise measurements of the number of lanes in a given base line has been

transported in all cases to date. Several types of aircraft have been used in this service. Those in the project field tests conducted in Louisiana are Canadian-made De Havilland "Beetles" and were assigned to the project by the U. S. Army 30th Engineers, Fort Stockton, California. A twin-engined Beech Bonanza and single-engined Beech Bonanza were used in a former USA-ERDL line measuring project conducted in Arizona.

The phase comparison receiver is a standard LoRAC positioning receiver converted to line measuring service. One LoRAC indicator indicates lanes and fractions down to 0.01 lane. The other indicator was modified to repeat the hundredths of a lane indication and gave in addition fractions of lane down to 0.001 lane.

The general mechanics of performing a line measurement has been described previously. The operation will now be described in somewhat greater detail.

The number of lanes contained in a line between the phase centers of the transmitting antennas may be found by dividing the line length by a lane width. Conversely, if a line is terminated by two transmitting antennas, the length of the line is determined if the number of lanes contained in the line can be found. The measuring by phase comparison consists in measuring the number of lanes contained in the unknown line.

If the base line determined by two LoRAC transmitting stations is extended beyond the transmitting station on each end of the base line, the extensions so generated are referred to simply as the base-line extensions.

Space surfaces having a constant phase comparison indicator reading are, to a high order of approximation, hyperboloids of revolution. The special hyperbolic sheet that passes through the mid-point of the base line is a plane perpendicular to the base line. Now if a LoRAC receiver starts from a point on a base-line extension and moves along any arbitrary path to a point on the opposite base-line extension, it must traverse every lane in the LoRAC system. It is only necessary for the receiver to produce a permanent record or count of the individual lanes traversed to know the total number of lanes in the base line.

In the field applications of the line measuring

technique described above, the phase comparison receiver is transported by means of aircraft in order to complete one traversal of the LoRAC system of hyperbolic lanes in the shortest possible time. At least five complete trips are usually run in order to ensure proper functioning of all equipment and to obtain a series of independent measurements. Since the aircraft cannot remain stationary on a base-line extension while a reading is being observed, the observed reading is obtained while the aircraft transits the base-line extension. A general description of the base-line crossing technique follows:

As the Lorac receiver leaves one base-line extension, the position indicator turns positively, or clockwise, and begins adding lanes; this base-line extension is called the minimum base-line extension. On leaving the opposite base-line extension, the position indicator will have reached a maximum count and will turn negatively, or subtract lanes. This extension is called the maximum base-line extension. As either base-line extension is approached, the position indicator approaches an extreme reading and reverses its direction of rotation as the extension is crossed. This reversal in the direction of rotation of the position indicator is easily readable to an observer, and the extreme meter reading is recorded. This operation is not difficult, since near the base-line extensions the lane expansion factor approaches infinity and the smallest graduations on the LoRAC position indicator (0.01 lane, for example) may occupy at least 0.25 mile on each side of the base-line extension. Thus for the last 0.01 lane to the extreme meter reading and back to the same meter reading will require 0.5 mile on the ground, which at a ground speed of 120 miles per hour gives 15 seconds or 0.25 minute for the observer to note the extreme indication.

Repeated crossings of a base-line extension provide a series of extreme position indicator readings. The observations so obtained are screened by a statistical method to eliminate errors and to give an average value calculated for the group.

The average minimum extension readings are subtracted from the average maximum extension readings, and the remainder is the uncorrected measured number of lanes in the base

TABLE 1. Computation of Line Length

Line no. 7	Date 3-31-56		Date 3-31-56		Date 3-31-56	
Measurement no.						
Maximum phase meter reading						
Minimum phase meter reading						
Measured lanes						
Propagation of phase correction in lanes						
Third frequency and reference phase correction in lanes						
Total corrected lanes						
Lane width, meters						
Semicorrected line length, meters						
Altitude correction, meters						
Line reduction to sea level, meters						
Total corrected distance, meters						

	7134 (1)	3889	1300 (2)	3801	7134 (3)	4048	1300 (4)	4151
	6215	7344	381	7282	6215	7344	381	7282
	918	6545	918	6519	918	6704	918	6869
	-	1104	-	1104	-	1104	-	1104
	+	0872	+	0872	+	0861	+	0862
	66.3618903	6313	918	6287	918	6461	918	6627
	60,962	110	66.3616253	694	66.3616328	855	66.3613594	706
	-	049	60,961	049	60,962	049	60,963	049
	-	082	-	082	-	082	-	082
	60,961	979	60,961	563	60,962	724	60,963	575

TABLE 2

Line	Measured Lanes	Reference Phase Cor- rection	Phase Cor- rection	Corrected Lanes	Lane Width	Semi- corrected Length	Alti- tude Cor- rection	Sea- Level Cor- rection	Line Length
1	1255.3939	-.0010	-.5984	1254.7945	66.3644156	83,273.704	+.043	-.049	83,273.698
2	1164.1293	+.0057	-.4639	1163.6711	66.3634127	77,225.185	+.052	-.048	77,225.189
3	2418.7879	+.1789	-.6133	2418.3535	66.3642696	160,492.264	-.231	-.048	160,491.985
4									
5	1419.2242	+.1152	-.3749	1418.9645	66.3634374	94,167.362	-.025	-.066	94,167.271
6	1188.5330	+.0845	-.4186	1188.1989	66.3630301	78,852.479	+.005	-.067	78,852.417
7	918.6545	+.0872	-.1104	918.6313	66.3618903	60,962.110	-.049	-.082	60,961.979
8	1650.0871	+.1173	-.3943	1649.8101	66.3615922	109,484.025	-.015	-.067	109,483.943
9									
10	1483.8769	+.1314	-.1658	1483.8425	66.3622262	98,471.092	-.046	-.081	98,470.965
11	1145.0349	+.1183	-.2641	1144.8891	66.3633125	75,978.633	+.019	-.066	75,977.586
12	1782.9793	+.1844	-.3958	1782.7679	66.3637712	118,311.201	-.095	-.065	118,311.041
13	2386.8318	+.2814	-.2579	2386.8553	66.3617902	158,395.991	-.061	-.082	158,395.848
14									
15	4031.5015	+.2832	-.3836	4031.4011	66.3604570	267,525.619	-.139	-.059	267,525.421
16	5032.5633	+.2704	-.5506	5032.2831	66.3594632	333,939.605	-.135	-.069	333,939.401
17	6233.4072	+.2683	-.6116	6233.0639	66.3603837	413,628.512	-.163	-.078	413,628.271

line. The corrections discussed previously are applied to the measured lane count, and the corrected measured lane count is then multiplied by the effective lane width to obtain line length. Examples of the results obtained during the Louisiana line measuring project are given below. Line numbers are identified in Figure 1, above. In Table 1 all computed corrections are given for all complete line measurements made on the sample line. This example shows the variation of the corrections throughout the period of time required to make a number of measurements.

In Table 2 the first measurement on each line is given except on lines on which a measurement was not completed. This example shows the variation of correction magnitude from one line to another. The effect of various types of terrain on correction magnitude can be seen in this table. Finally, in Table 3 the results of all line measurements are compared with the USC&GS length as standard. This table shows that a high degree of accuracy was achieved in all cases.

LINE MEASURING USING MULTIPLE FREQUENCIES

A variation of the line measuring techniques described previously was partly tested during

project field operations. For the purposes of describing this method suppose the following: (1) A given line is to be measured whose length is known approximately. The accuracy to which the length is known need not be any better than that obtained by inverse computation from astronomic positions at the ends of the line. (2) A multiplicity of measuring frequencies is available where the multiplicity is not less than 2 and preferably 3 or more. If the given line is measured using one of the frequencies from (2) above, as previously described, an integral number of lanes plus some fraction of a lane will result. Each time the line measurement is performed using a different frequency from (2), a new integral number and fraction are obtained.

Consider a nomogram constructed for a given line measurement as follows. Starting from the left-hand end of a straight line as the origin, lay out a distance to scale which represents the fraction of a lane obtained in the line measurement. If the fraction of a lane is multiplied by a lane width, the scale on the line will be directly in feet or meters. From the terminal point of the scaled length representing the measured fraction of a lane, lay out a series of distances to scale, each equal to an integral lane width. Continue this series of line segments

TABLE 3

Line No.	Number of Measurements	Geodetic Length	Electronic Length	Difference	Relative Accuracy
1	3	83,272.604	82273.520	+0.916	1 : 90,900
2	5	77,226.033	77225.481	-0.552	1 : 139,900
3	4	160,498.074	160493.469	-4.605	1 : 34,850
4		(68,453.275)		No measurement	
5	4	94,165.810	94166.894	+1.084	1 : 87,790
6	4	78,852.410	78852.760	+0.350	1 : 225,290
7	4	60,960.493	60962.460	+1.967	1 : 30,990
8	4	109,484.039	109484.305	+0.266	1 : 411,590
9		(114,849.546)		No measurement	
10	4	98,472.960	98470.812	-2.148	1 : 45,840
11	4	75,975.358	75977.378	+2.020	1 : 37,610
12	4	118,307.045	118309.749	+2.704	1 : 47,750
13	4	158,395.545	158394.626	-0.919	1 : 172,360
15*	3	267,525.000	267520.914	-4.086	1 : 65,470
16	1	333,947.945	333939.401	-8.544	1 : 39,090
17	2	413,629.482	413629.784	+0.302	1 : 1,369,630
		2,130,712.798	2,130,701.553		

All lengths in meters.

* Line 14 was deleted from the project.

until a number is obtained equal to the measured number of integral lanes in the line.

The scaled fraction plus the measured integral lanes is now a graphical representation of the line length measurement. To improve the appearance of the graphic plot, short vertical risers of uniform length are constructed at the origin and terminus of each line segment on the plot. The base of each vertical riser is connected by a straight line to the top of the next riser on its right. This construction creates a sawtoothed plot, in which each sawtooth represents one LoRAC lane. By constructing similar plots for the remaining independent line measurements on line bases which are parallel to one another, and with vertical alignment of the origins, the line measurements can be compared. In this respect, since the same line is being measured each time, the vertical risers for all plots will be in vertical alignment at the right-hand end of the line plots. Since each line measurement is made on a different frequency, individual sawteeth in each measurement plot will have a different width. If the vertical alignment of the risers is compared from the end of the line plot toward the origin, a progressive misalignment will be noted. After

a certain distance along the graphical plot from the end of the line toward the origin, a second vertical alignment may be noted.

It can be seen at this point that all that is actually necessary to plot the nomograms described above is (1) the fraction of a lane in a line measurement, and (2) the frequency used during the measurement. Of these two requirements, only (1) must be measured in the field; the frequency used to perform the measurement is an instrumentation factor.

The integral lane count may be determined by the vertical alignment characteristic of the graphical plots. The correct vertical alignment may be selected out of all such alignments that occur by the known approximate length of the line. This information suggests the following modification in the line measuring technique. Instead of one LoRAC receiver performing the line measurement, two would be used. They would be checked against each other at a common point to be sure that they indicate identically when identical points are occupied. A line measurement would begin by having the two receivers on station near the base-line crossing point and in communication with each other. At a present time both receivers would

in a series of base-line extension crossings. After a prescribed number of crossings had been made, the data so collected would be considered as one line measurement. The procedure would be repeated several times to obtain the best possible measurement. After a sufficient number of line measurements had been collected, the base-line transmitting stations would change to a new measuring frequency and the line measuring procedure would be repeated. The line measurement would be complete when all allotted measuring frequencies had been used. The basic line measurement data would be processed as before, and a chart

would be constructed to permit the selection of the proper lane coincidences.

CONCLUSIONS

The experimental results obtained during the two line measuring projects show that good accuracy in line length results can be obtained by phase comparison measurements. The phase delays along all types of lines were predicted with better than expected accuracy, and homogeneous transmission path lines can be predicted with excellent results. Further line measuring projects would make further gains in accuracy through better computational techniques and field practice.

Microwave Position-Fixing System

H. R. SMYTH

*National Research Council of Canada
Ottawa, Canada*

Introduction. Shortly after merchant marine radar came into use as a navigational aid for vessels, it was thought that some consideration should be given to the development of electronic aids for vessels too small to use radar, such as tugs and fishing vessels. It was obvious that the use of the microwave frequencies of radar would best suit the development of a small direction-finding receiver. The final system developed by the National Research Council used a compact 3-cm pulse transmitter on a shore site, such as a lighthouse or harbor breakwater. The antenna was designed to radiate energy across the water from all angles of approach.

The receiving equipment was built into a case on the back of a small radar antenna consisting of a 6-in. slotted wave guide at the focal plane of a parabolic cylinder. The receiver used a radar crystal detector and transistorized audio amplifier for reception on phones or a speaker. The receiver could be hand held or mounted on a pelorus head, and rotated around the horizon to take a bearing on the shore transmitter. The signal from the shore is audible over an arc of only about 6° and can be maximized with an accuracy of $\pm 1^\circ$.

The over-all system was very successful, and consistent range was obtained to the horizon. A high percentage of the time, the range was many times the horizon range, the refractive index of the atmosphere being such as to bend the signal to the earth's curvature in a trapping action in the troposphere.

Soon after completion of the development of this direction finder, work was started on a visual display system for the receiver. It was believed that a visual display could be used to view the bearing of three shore transmitters simultaneously, and obtain an instantaneous fix,

rather than a running fix from one shore station.

It was suggested that, by displaying bearings of the three transmitters on a large cathode-ray tube, the angles could be measured more quickly and accurately. To test this idea an experimental set was built with a circular sweep on a 10-in. cathode-ray tube. The electron beam of the tube was made to rotate synchronously with a radar-type antenna. The speed of rotation was 22 rpm. A simple crystal video receiver amplified the signals received and its output brightened the cathode-ray-tube display on each bearing of the shore transmitters. The three short, bright arcs formed represented the bearings of the three transmitters. Each of these arcs was bisected by the cursor arms of a three-station pointer, whose center coincided with the center of rotation of the cathode-ray display. Trials of this type of equipment in the field indicated several shortcomings. (1) Bisection of the arc with the three-station pointer caused severe eyestrain on the operator, who was required to work in semi-darkness. (2) The earth's field acting on the cathode-ray beam caused decentering of the sweep, and it became necessary to add accurate centering controls, for adjustment every time

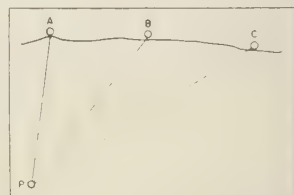
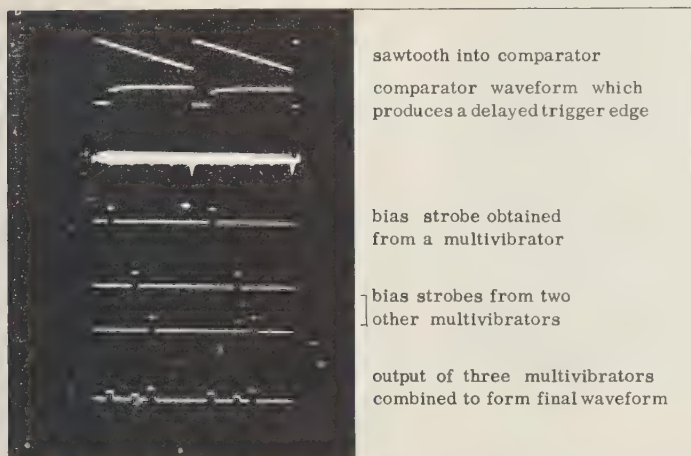


Fig. 1. Position of ship, *P*, with respect to three shore stations, *A*, *B*, and *C*.



2. Wave forms showing the generation of 'bias strobes' (length of sweep represents two rotations of antenna).

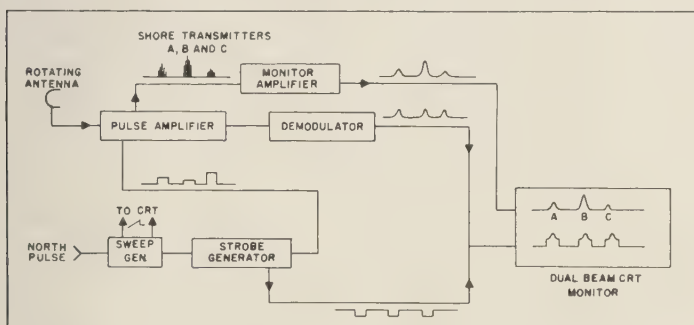


Fig. 3a. Block diagram of connections to the dual-beam cathode-ray-tube monitor.

ship altered course and change the angle of earth's field in relation to the cathode-ray e. (3) Difficulties with the mechanical de- of the rotating magnetic field also ap- as small errors in bearings. This com- of faults inherent in the system ulted in a piece of equipment with angular uracy of about $\pm 0.5^\circ$, and in some cases h operator fatigue it amounted to $\pm 0.75^\circ$. was therefore decided to discard the system favor of one that would eliminate the in- erent faults.

Electronic angle measurement with digital play. This system retains the identical re transmitter stations that radiate a signal the ship desiring to fix its position accurately

for hydrographic work. The receiving equip- ment on the ship uses the same type of re- ceiving equipment, but the method of indicat- ing the bearings of the short stations eliminates the difficulties due to operator fatigue, earth's field, display distortion, and centering difficul- ties with the rotating magnetic field.

Basically, this method involves finding the center of the receiving antenna pattern for each of the three shore stations by electronic means rather than by bisecting the brightened arc manually, as was done before. Two coun- ters are used to record the angles, the first counter being 'opened' at the maximum signal from the first station and 'closed' at the maxi- mum of the second station. The second counter



Fig. 3b. A single shore transmitter signal as it appears on the monitor display: *above*, with bias strobe adjusted to pass a signal through the receiver; *below*, with bias strobe adjusted to cut off receiver output.

is 'opened' simultaneously with the 'closing' of the first counter, and closed at the maximum of the third station. During the time that each counter is 'opened,' finely divided angular markers (0.02° apart) enter the counters, and the angles subtended between shore stations are displayed after one revolution of the receiving antenna. The angles displayed on the counters are stored until erased electrically before the next set of angles is measured.

Antenna rotation speed at the ship was in-

creased to 1 rps, so that the angles could be measured as rapidly as possible before the speed of the vessel past the shore stations could alter the true angle between stations, owing to the relative velocities of the ship to the shore velocity of the rotating antenna. Errors in measurement with a slowly rotating antenna are positive when the shore stations and antenna motion are clockwise and negative when the relative motion of stations and antenna are opposite. The resulting errors decrease with increasing range from the shore stations. The errors also decrease in proportion to the increasing speed of rotation of the antenna. At 1 rps errors are negligible.

In order to be sure that the measured angles will always be started from a known station, it was necessary to provide a switching pulse when the rotating antenna on the ship passed through a north bearing. Then the first station scanned by the antenna after the 'north' switching pulse will initiate the angle counter between station A and B, and then the second counter will operate to measure the angle between B and C (Fig. 1).

Receiver gain control and monitor. Signals received by the rotating antenna from each of the three shore stations oscillate about the free space value as the ship moves closer to or farther away from the stations. A method of controlling the gain of the receiver so that the signals do not become distorted or saturated will now be described.

A dual-beam cathode-ray tube monitors the 'raw' signals and the 'controlled' signals. The main receiving antenna feeds the input of an ordinary crystal video amplifier, but the strength of the signals received from each of the three shore stations varies with the distance involved. To measure the subtended angles between the stations accurately, the signals must be approximately equal in strength, so that the distortion of a strong signal in a saturated receiver cannot give a false electrical center of the received signal.

To overcome this variation of signal strength, three 'bias strobes' were generated to handle the individual gain of each received signal. Each of the three strobes was separately controlled in azimuth, so that an operator could move the strobe to coincide with each signal

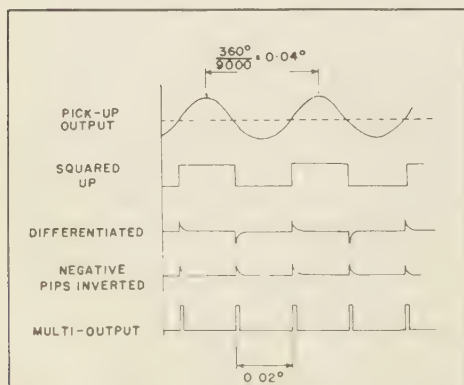


Fig. 4. Generation of 0.02° angular markers.

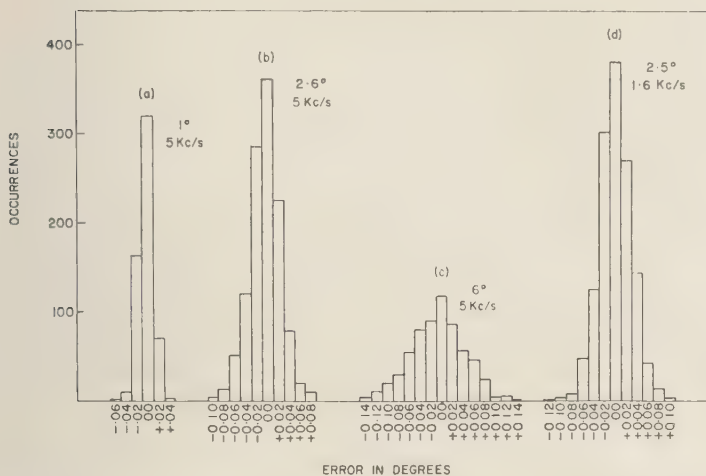


Fig. 5. Distribution of error on measurements taken on an angle of 360° for different antenna beam-widths and pulse repetition frequencies. Antenna speed = 58.8 rpm.

three gain controls riding the strobes could be adjusted to equalize the receiver signals (Fig. 2).

The three signals appear on Figure 3a. The monitor scope displays the three shore transmitter signals, each riding its respective strobe pedestal. They have the appearance of a Loran signal riding its pedestal.

Movement of the vessel during a survey operation changes the bearing of each shore station slowly, and the bias strobes are moved by the manual controls to maintain the signals approximately centered. This, and the manipulation of the three gain controls, do not occur rapidly, and an operator has no difficulty maintaining the adjustment. The controls require adjustment only at spaced intervals, depending on the distance of the ship offshore and the speed and direction of the operation.

Monitor display. The display cathode-ray tube provides two A-scan sweeps, one above the other. The sweep is initiated from the left by the north pulse, and the linear sweep represents 360° of antenna rotation over to the right-hand side.

The positions of the three received signals are monitored on the top trace of the monitor cathode-ray tube. The three strobes and their associated gain controls previously mentioned isolate the individual three signals and display

them with the strobe pedestals on the lower monitor trace. Figure 3b shows a single shore transmitter signal as it appears on the monitor display.

The use of three gain-controlled pedestals serves another useful purpose. Since the receiver gain is turned on only during the instant when the antenna is sweeping past a shore station, it follows that no signals of any kind are received while the antenna swings through the remaining bearings in 360° . This automatically eliminates interference from marine or airborne radar equipment in the same area. Further reduction in the possibility of interference from other radar sets that might be on the same bearing as a shore station is achieved by cross polarization. The position-fixing equipment uses microwave antennas designed for vertically polarized waves; marine radar apparatus uses horizontally polarized waves. Possible interference is thereby reduced considerably, as only the rotated reflections of the marine radar are effective in producing a vertically polarized objectionable signal.

Angular markers and counter display. A gear train increases the speed of a slotted tone wheel to 30 times the speed of the antenna. The tone wheel has 300 slots. Thus one revolution of the antenna is divided into 9000 parts. A simple system of light beam and phototran-

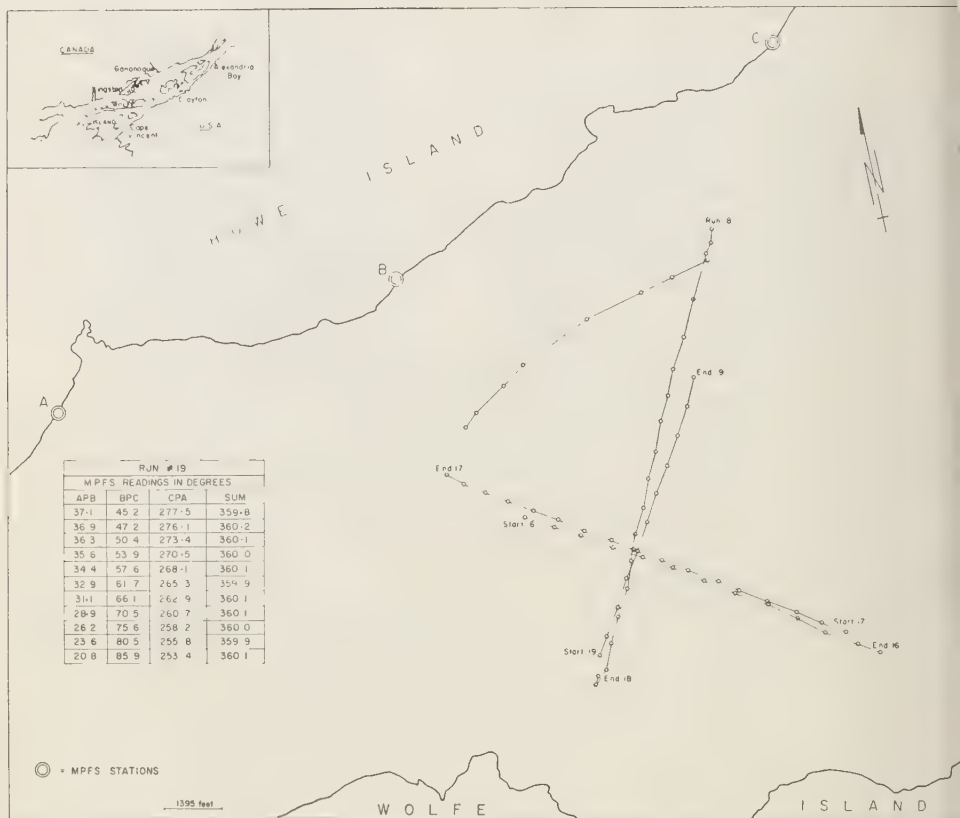


Fig. 6. Insert at top left shows location of test site. Fixes taken from angles measured electronically. Table shows counter readings for run 19, *M. V. Radel II*, July 1957.

sistor, interrupted by the tone wheel, produces 9000 cycles, representing 0.04° for each cycle, and from Figure 4 it will be seen that two pulses are formed from each cycle for 0.02° measurements.

These pulses are recorded on the counter display only when the gate signifying the center line to a shore transmitter has been passed. The north pulse, previously discussed, orients the gates so that the two angles are registered in the same sequence during each sweep of the antenna. The north pulse position is stabilized through a synchro differential whose shaft is controlled by servo drive. The servo input is a magnetic compass which has a synchro pick-off. Thus, altering the ship's course does not affect the position of the north pulse on the

previously adjusted gain pedestals straddling each of the shore stations.

Laboratory tests. A simple method of checking the accuracy of the equipment consists in taking measurements on a single transmitter i.e., on an angle of 360° . Several different antennas were tried, their beamwidths varying from 6° to 1° . Figure 5 shows a frequency distribution of the errors; (a), (b), and (c) show the distribution at a pulse repetition frequency of 5000 pps, and (d) shows the distribution for a 2.5° antenna at 1600 pps.

Field tests in 1957. Initial field testing was carried out from the *Radel II* on Lake Ontario during July 1957 and in L'Orignal Bay on the Ottawa River in October of the same year. The main object of the Lake Ontario tests was to

TABLE 1. Comparison of Sextant Angles and MPFS Angles

'Diff.' equals 'MPFS' minus 'Sextant'. Readings taken from *M. V. Radel II*, October 1957, in L'Original Bay on the Ottawa River.

Ship at Anchor				Ship at Half Speed (7 knots)				Ship at Full Speed (11 knots)			
Sex-	Diff.	MPFS	Sex-	Diff.	MPFS	Sex-	Diff.	MPFS	Sex-	Diff.	MPFS
tant			tant			tant			tant		
Stabilized by Gyro											
4°55'	-13'	34°42'	53°02'	-20'	52°42'	36°42'	-12'	36°30'	44°54'	00'	44°54'
4°56	-08	34 48	53 04	-04	53 00	37 56	-02	37 54	45 05	-05	45 00
4°54	-06	34 48	53 00	-06	52 52	39 05	-17	38 48	45 09	-09	45 00
4°53	-05	34 48	53 00	-06	52 54	40 19	-07	40 12	45 06	-06	45 00
4°52	-16	34 36	52 59	+07	53 06	41 47	+01	41 48	45 00	-12	44 48
4°51	-09	34 42	52 58	-04	52 54	43 10	-04	43 06	44 50	-08	44 42
4°50	-08	34 42	52 58	-04	52 54	44 32	+04	44 36	44 33	-15	44 18
4°49	-07	34 42	53 00	00	53 00	46 06	00	46 06	44 06	-12	43 54
4°50	-14	34 36	53 00	00	53 00	47 37	+05	47 42	43 40	-10	43 30
						49 13	-01	49 12	43 00	-06	42 54
									62 37	+17	62 54
									33 37	-25	33 12
Unstabilized											
34°50'	-09'	34°41'	52°57'	-04'	52°53'	24°05'	-04'	24°01'	35°12'	-20'	34°52'
34°49	-07	34 42	52 57	-01	52 56	25 39	-01	25 38	37 35	-11	37 24
34°49	-07	34 42	52 58	-01	52 57	26 54	-04	26 50	39 00	-07	38 53
34°49	-07	34 42	52 58	-02	52 56	27 51	-03	27 48	39 58	-04	39 54
34°47	-07	34 40	52 58	-03	52 55	28 51	-01	28 50	40 56	-08	40 48
34°47	-07	34 40	52 57	-04	52 53	29 55	-03	29 52	41 47	-10	41 37
34°48	-08	34 40	52 58	-07	52 51	30 54	-03	30 51	42 29	-07	42 22
34°47	-07	34 40	52 58	-05	52 53	31 55	-02	31 53	43 05	-09	42 56
34°46	-06	34 40	52 58	-04	52 54	32 54	-02	32 52	43 40	-05	43 35
34°47	-09	34 38	52 57	-07	52 50	34 15	-05	34 10	44 18	-10	44 08
									62 05	+04	62 09
									17 13	-03	17 10

find out whether this system was at all practical. The purpose of the tests on the Ottawa River was to get some idea of the angular accuracy of the bearing measurements.

With respect to the first test (Fig. 6) three shore transmitters were set up on Howe Island and the receiving equipment on the *Radel II*. The two included angles as well as the closing angle were measured. (Three sets of counters were used in the experimental set.) The three readings were totaled to check against 360°. The included angles were set on a three-station pointer and 'plotted' on a scale of 1 in. = 1395 ft. 'Fixes' were taken at a speed of about 7 to 8 knots, and no difficulties were met in keeping track of the positions and amplitudes of the manually operated bias strobes. However, it became evident that a faster scanning speed would reduce the distance the ship traveled during an angle measurement. The antenna rotated at 22 rpm, or about 1 revolution in 3 seconds.

In the latter part of July, the system was tested in the open lake for maximum range.

Three stations were set up: *A*, at the entrance to Toronto Harbour; *B*, at Scarboro; and *C*, at the mouth of the Rouge River. The separation between stations was about 7 miles. Strong signals were received at distances far exceeding line of sight because of 'anomalous' propagation conditions, and no conclusive figures for maximum range were obtained in the field. (As an example, fixes were taken beyond 20 miles with transmitting and receiving antenna heights of about 20 ft above the water line. Under 'standard' propagation conditions, the expected range for these heights would be about 10 miles.)

With the assistance of the Canadian Hydrographic Service, comparisons were made of the angles measured electronically and visually. This test was carried out in L'Original Bay on the Ottawa River. The three stations were placed about 1 mile apart, and readings were taken from distances varying between 5000 and 9000 ft. At these distances, visual markers on top of each shore transmitter were clearly visible. The results are shown in Table 1. It should

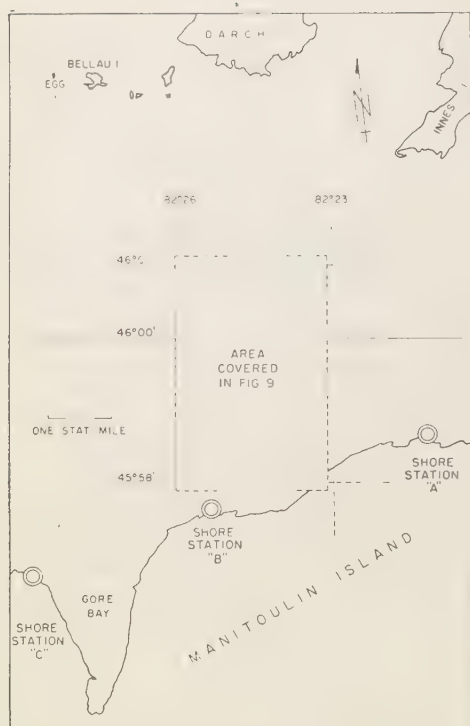


Fig. 7. Chart showing test area in north channel of Lake Huron.

be noted that in one case (unstabilized) readings were taken with an angular marker spacing of 0.01° . These were generated from a high-speed toothed wheel attached to the antenna motor shaft. In the other case (stabilized) readings were taken with an angular marker spacing of 0.1° . These were generated by another toothed wheel which was driven through a mechanical differential. The ship's course was fed into this differential through a Sperry repeater motor and gyro, so that any change in course during an angle measurement would not affect the number of angular markers generated by the toothed wheel.

Field tests in 1958. A new set was built in 1958 to be used on the survey launch *Boulton* (Canadian Hydrographic Service). The antenna speed was increased to approximately 60 rpm from the original 22. The angular marker spacing was set to 0.02° with no stabilization. Also,

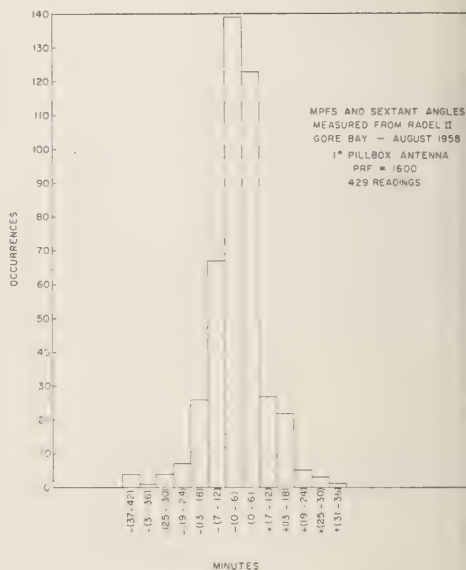


Fig. 8. Comparison of MPFS angles and sextant angles.

two sets of counters rather than three sets were used since it was considered unnecessary to read out the back angle.

The main reason for increasing antenna rotational speed was to reduce the time required during angle measurement. This reduces the distance the ship travels during the measurement. As an example, for a ship's speed of 10 knots (10×2000 yd/hr) the distance traveled in 1 second is about 50 ft. The antenna scans 360° /second, which means that, if $APB + BPC = 180^\circ$, the ship has moved 25 ft. Normally, the angles total to much less than 180° . Also, by reducing the time required for the angle measurement, it was thought that the angular markers (which are generated by the toothed wheel geared directly to the antenna) need not be stabilized against turning of the ship.

In August 1958 the equipment was taken to an area in the North Channel of Lake Huron on the *Radel II* (Fig. 7). A series of angle measurements were made on the three shore stations which were placed directly above the hydrographic survey markers. Sextant readings were taken simultaneously. The differences be-

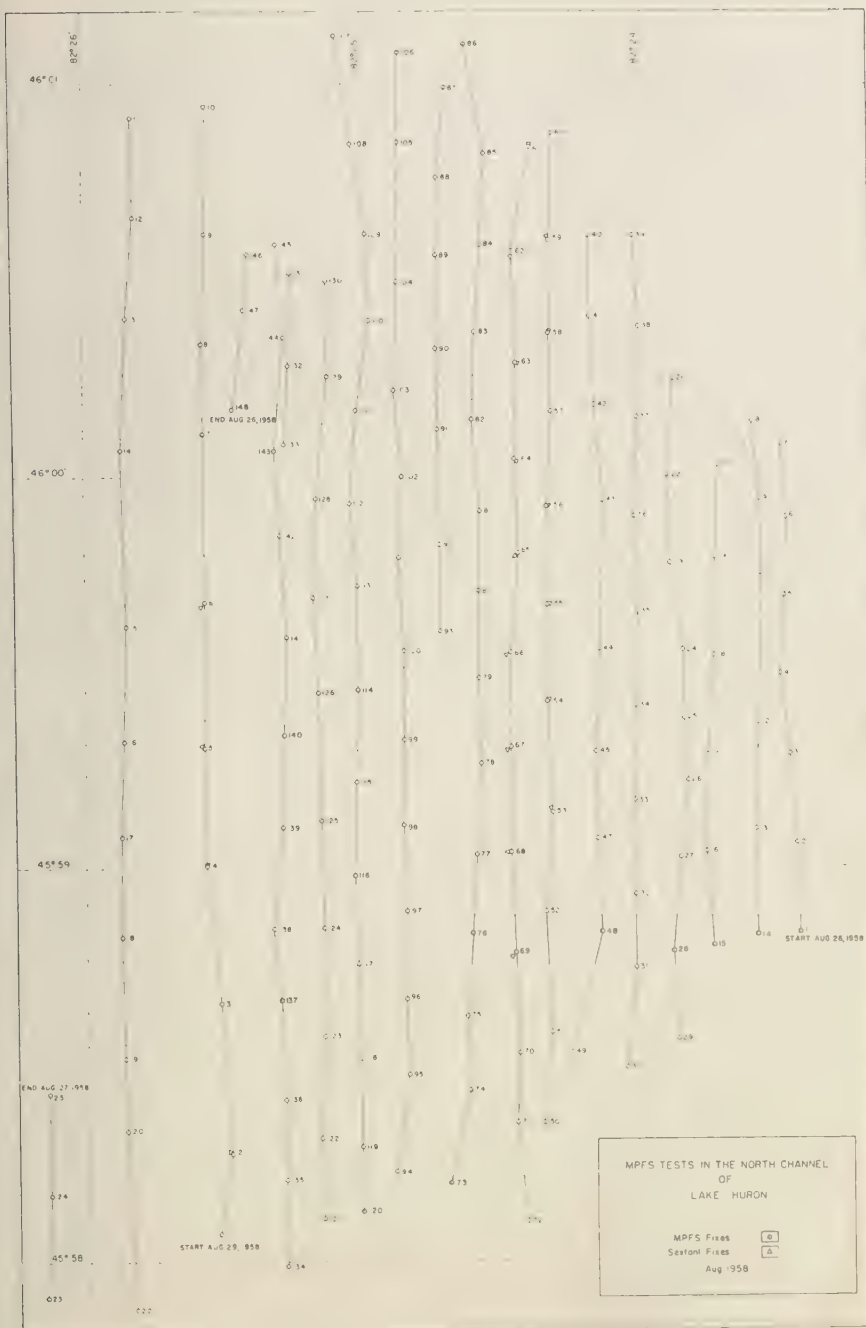


Fig. 9. 'Sounding lines' run with MPFS, (numbers indicate fixes).

tween electronic and sextant readings were plotted in Figure 8. Of the 429 readings, 262 were within the range ± 6 minutes, 356 readings within ± 12 minutes, and 404 within ± 18 minutes. The accuracy of the sextant readings was estimated to be within ± 5 minutes.

The equipment was transferred from the *Radel II* to the *Boulton* in the latter part of August. 'Lines' were run about 600 ft apart on a scale of 1 in. = 1000 ft. Fixes were taken about 2 minutes apart as a check. The results are shown in Figure 9.

Electronic Control Systems Used on Hydrographic Surveys

GILBERT R. FISH

*U. S. Coast and Geodetic Survey
Washington, D. C.*

During the years 1957 and 1958 I had command of the U. S. Coast and Geodetic Survey ship *Hydrographer* while it was engaged on a survey of Georges Bank, which lies about 100 miles east of Cape Cod on the Massachusetts coast. (See Fig. 1.) This was a very interesting survey in that three electronic control systems were used to measure distance and furnish position control for the sounding lines. The principal system used was Raydist, and all the sounding lines run by the ship on the Bank were controlled by this system. In some of the deeper water around the edges of the Bank the electronic position indicator was used to control the sounding lines; and in the shoal areas on the Bank, which were too shallow for the ship, Shoran was used to control the position of the sounding launch.

To provide control in the offshore areas beyond the range of Shoran, the Coast and Geodetic Survey developed a system that we call the electronic position indicator. Shoran uses high-frequency radio transmitters, and the radio waves tend to shoot off into space rather than bend over the horizon. This limits the range to a distance only slightly greater than the line-of-sight distance between the antenna on the ship and the antenna on the shore station. The electronic position indicator uses radio waves of a lower frequency and longer wavelength, and these waves will bend over the horizon. This permits the EPI to measure distance when the ship is as much as 450 miles from the shore station.

Shoran will measure distance with sufficient accuracy so that it can be used to control surveys at scales as large as 1 : 10,000, if sufficient precautions are taken with the calibration measurements. The EPI, however, owing to the use of radio waves with a longer wave-

length and to the fact that the personal factor in reading the distance has greater effect than with Shoran, is not suitable for use as control on large-scale surveys and is best confined to surveys at scales of 1 : 100,000 or smaller.

In planning for the survey on Georges Bank it was decided that the principal scale to be used to develop the area would be 1 : 40,000, with larger scales in some of the shoaler areas. Neither Shoran nor the EPI was quite suitable as a control method, the first because of the distance from land, even though one fixed station might be placed on the Texas tower located on the shoal, and the second because it did not have the necessary accuracy. In 1954 the Coast and Geodetic Survey had successfully used a Raydist system on a trial basis to make an inshore survey on the west coast of Florida. It was decided that Raydist would work equally as well on Georges Bank if a set could be obtained that had sufficient power to permit strong radio signals to be received on the Bank. An arrangement was therefore made with the Hastings-Raydist Company of Hampton, Virginia, for the trial of a Raydist system, with an agreement to purchase the set if it proved satisfactory and funds were available. The Raydist system worked in a very satisfactory manner and was purchased. During the 1957 and 1958 survey seasons about 17,000 miles of hydrography were accomplished using Raydist for control.

Principles of operation. The Raydist set purchased by the Coast and Geodetic Survey is a DM model with 100 watts of power. It is a distance-elliptical system, where the distance to each of two shore stations appears on the phase meter dials. In some of the earlier models the distance to the second shore station was not read direct but had to be obtained by a

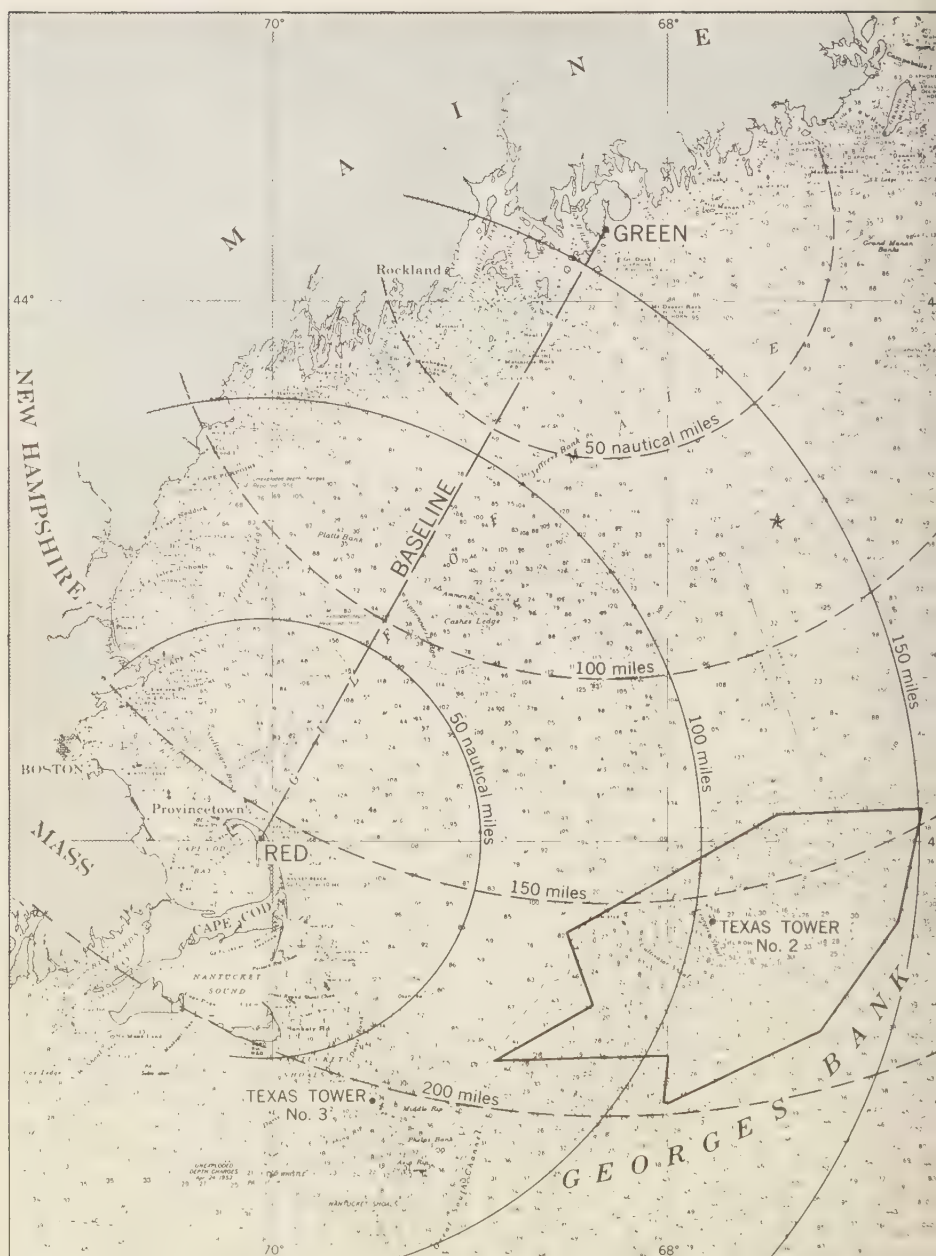


Fig. 1. Project area of 1957 and 1958 Raydist-controlled survey of Georges Bank.

rt computation, using the dial reading and base-line distance between the two shore stations.

Raydist is a phase-measuring instrument, and the determination of distance is based on measurement of a phase difference in two radio signals, one emanating from a transmitter aboard ship and the other from a transmitter one of the fixed, or shore, stations. In order to measure the distance of the ship from each of two shore stations it is necessary to measure the phase difference at three places, aboard ship and at each shore station. The change in phase relationship as the ship changes distance from a shore station causes the dials to turn, and there is no necessity to match pipes in a pulsing-type system.

Four radio transmitters and frequencies are required to operate the system, the two distance-measuring frequencies mentioned above and two additional frequencies for a transmitter at each shore station to transmit the information about the phase relationship at the Red station back to the ship. In referring to the shore stations, the station with the distance-measuring transmitter is called the R_1 , or 'Red,' station, and other is called the R_2 , or 'Green,' station. The frequencies used on the *Hydrographer* were 3280 kc/s aboard ship for the distance-measuring transmitter, 1639.8 kc/s at the Red station for the transmitter whose frequency is compared with the 3280-kc/s transmitter, and 2492 and 2496 kc/s for the carrier frequencies. The frequency of the 1639.8-kc/s signal is doubled to 3279.6 kc/s before comparison with the 3280 signal so that it will be 400 cycles per second less than the latter frequency. Because it is difficult to measure phase directly at the radio frequencies a technique known as heterodyning is used. In this technique, if two radio signals are fed into a detector the resulting signal will have a frequency equal to the difference between the frequencies of the entering signals, and will portray the phase changes that take place between the entering signals. In the Raydist the entering signals are separated by 400 cps, and after detection the frequency is fed into a phase meter and used to turn the dials. It is easier to handle the phase measurement at 400 cps than at the higher radio frequencies.

The distance from the ship to the Red station is measured by comparing the phase of the transmitter on the ship to the transmitter at the Red station. Distance is measured by knowing how many half-wavelengths of transmitter frequency there are between the shore station and the ship. Once this number is known and the phase meter dial set to this value, at a point where the distance to the shore station is known, the Raydist system keeps count of the number of half-wavelengths as the ship moves. These half-wavelengths are known as lanes, and for the radio frequencies used in the Raydist aboard the *Hydrographer* they are about 45 meters, or 150 ft, wide.

The ship and the Red station transmitters are crystal controlled, but there is a small change in phase with respect to each other since they are not phase locked. To overcome this the phase of the two transmitters is compared at the Red station and the phase information in the 400-cps difference frequency is relayed back to the ship on one of the carrier frequencies. At the same instant the two transmitter frequencies are compared on the ship and the phase information appears in this 400-cps difference in their frequencies. These two 400-cps signals represent the difference in phase at each end of the R_1 line being measured. The arbitrary differences between the two transmitters, which may be the result of small fluctuations in frequency, are equal at each end of the line and hence cancel. The phase differences that do appear, and are used to turn the phase meter dials, are the result of changes in phase caused by a change in the R_1 distance between the ship and the shore station.

In the distance-elliptical Raydist system the R_2 distance is not measured direct but is derived from knowledge of the R_1 distance and the sum of the R_1 and R_2 distances. This is done to simplify the system and to prevent having to use additional radio frequencies. The phase of the ship and the Red station transmitters is compared at the Green station, and the information is relayed back to the ship on the carrier frequency at that station. At the same instant the ship and the Red station transmitters are compared on the ship. The two 400-cps signals are put into an elliptical phase resolver, the shaft position of which is related

to the sum distances R_1 and R_2 . By combining the distance information from the Red station phase meter with the elliptical resolver, and by a proper gearing arrangement, the Green phase meter dial reads only the value R_2 , or a direct distance to the Green station.

To furnish additional information on the correctness of the lane count as shown on the phase meter dials a brush recorder makes a record of the change in lane count on a special tape. This tape is monitored, and the hydrographer has available information that will enable him to stop the hydrography when the lane count is incorrect.

As was mentioned previously, it is necessary to set the Raydist phase meter dials at a known point, but no additional calibration is necessary. With a pulsing-type system the rapidity with which the leading edge of the pulse builds up is a function of the signal strength and varies with the distance of the ship from the shore station. Raydist being a phase-measuring system is not subject to this type of error. Also small changes in the characteristics of the electronic circuits do not affect the distance as much as they do in a pulsing system. All types of systems, though, are subject to errors when the velocity of propagation of the radio wave, which is considered to proceed over water, is changed by intervening land between the ship and the shore station. For the Georges Bank project the velocity of propagation was taken as 938,167,315 ft/s.

At the Red station three antennas are required, one receiving and two transmitting. In 1957 all three were 100-ft towers with ground planes. In 1958 a 35-ft whip antenna was used for the receiving antenna, and it was duplexed by placing it on top of the 100-ft tower antenna used to transmit the 1639.8-ke/s signal. This reduced the requirement for 100-ft tower antennas at the Red station to two. At the Green station two 100-ft tower antennas with ground planes are required.

On Georges Bank the distance to the Red station on Cape Cod was about 100 miles, and the distance to the Green station at Southwest Harbor, Maine, was about 150 miles. At the end of the 1958 field season the latter distance had increased to 175 miles. During the daylight hours strong signals were received and the phase meter dials were ordinarily quite steady

and well locked into the correct lane count. With the coming of darkness the sky waves would begin to be received, and they would affect the stability of the phase meter dial indicating distance from the station in Maine. On an ordinary day the lane count would be lost several hours after darkness began.

In the summer of 1958 the ship worked the special project southeast of Georges Bank. Raydist was not used for control, but it was turned on and a strong signal was received from the station in Maine when the ship was 225 miles from the station. At the same time the signal from the station on Cape Cod was weak, even though the distance to this station was only 175 miles. This was due to the complex antenna used at this station in 1958, as mentioned previously. If the duplexed antenna is not used, the signal from this station should have as much range as, or more range than, that from the other station. No attempt was made to determine the maximum range of the Raydist, but it may be as great as 200 miles under good conditions. During the use of the Raydist, distance from the shore station seemed to have an adverse effect if the radio propagation conditions were poor.

Radio interference. With Raydist the signals emitted by the four radio transmitters must be received continuously, and interference from any of these frequencies will affect the operation of the lane counters in the phase meter. On Georges Bank we experienced some interference from other users of the 2492 and 1639.8 ke/s we were using as carrier frequencies. Raydist has a Monitor that can be used to tune in to the tone of the 400-cps difference frequency signals as originated aboard ship or either of the shore stations, and the radio interference would be easily recognizable in the signal. The Boston Marine Operator broadcasts on a frequency of 2505 ke/s, and if it would be almost possible to make out the words in the message. Interference of this type did not appear to affect the operation of Raydist adversely if the signals were continuous in strong, but at night when the sky waves were present this type of interference was often the deciding factor as to whether the phase meter dials would continue to operate with the correct lane count.

Static conditions will affect the operation

phase meter dials, and its effect is worse at night when the sky waves are present. The most kind was precipitation static. With static present, if rain began to fall on the ship it was almost certain that the lane count would be lost. A steady fall of rain did not have any adverse effect. Static or thundershower conditions along any of the paths followed by the radio signals, or near either the ship or the radio stations, may cause trouble. Some time was lost on Georges Bank on account of this condition, but on even the worst days it was possible to accomplish some hydrography.

As was mentioned previously, Raydist is a distance-measuring system, and the factor that determined the length of the ordinary work day on Georges Bank was the presence of sky waves. The signal received from the Green station during the hours of darkness. During the day, the hours the phase meter dials would be automatically locked to the lane count, but shortly after sunset it was normal to begin to receive the sky waves. With the sky waves arriving out of phase with the ground waves the phase meter dials begin to oscillate through an arc that increases in size as the strength of the sky wave increases relative to that of the ground wave. Sometimes the Green station dial would oscillate through an arc that might approach a complete revolution and still return to the correct lane count. At times extra lanes would be put in and then a few minutes later be taken out again. If it was possible to keep track of the lane count so that the error did not exceed one lane, hydrography would continue if being run at a scale of 1 : 40,000. When there was a probability that the error might exceed this amount hydrography was terminated.

Often, after a period of sky-wave reception immediately after sunset, the Raydist dials would settle down and be as steady as during the daylight hours. This condition usually would last more than a few hours, and it was possible that lane count would remain accurate until after midnight. There were four nights each year when the Raydist worked all night, and on only three of the eight nights was there no gain or loss of lanes. Interestingly, one of nights when no loss of lanes occurred was a night when the northern lights were very prominent and ordinary radio recep-

tion was poor. Operations in the morning usually began at sunrise. On some mornings the sky waves would affect the Green station dial right up to the time the sun was visible, and there might be an especially bad burst of activity a few minutes before sunrise. With the appearance of the sun the Raydist dials would usually settle down and remain stable all day.

Use of Raydist. Because the Raydist dials must be set at a known point it is necessary to have some kind of a reference point in the survey area. On Georges Bank we were fortunate in that the U. S. Air Force has built a Texas tower, and we used one leg of it for a permanent reference point. The position of the tower was determined by setting the Raydist dials at Cape Cod and then making the run to Georges Bank. The agreement between the runs was good. One run was made between the Texas tower on Georges Bank and the Texas tower on Nantucket Shoals, about 70 miles southeast of the Red station on Cape Cod. The position of the tower on Georges Bank as determined by this run was in good agreement with the values obtained by reference to Cape Cod, indicating good distance-measuring qualities for the Raydist.

When working in areas distant from the tower on Georges Bank it was necessary to have a reference point for setting the dials in the vicinity of the work. Buoys were used for these points, and by planting the buoys in the more shallow water with a short scope of anchor line, and by considering the scope of the anchor line, it was possible to set the dials accurately with reference to the position of the tower. One problem with the buoys was finding them in the fog that is so prevalent on Georges Bank during the summer months. To assist in the search, a 2 by 2 in. pole was placed on top of the buoy and 1-lb coffee cans were nailed to the four sides of the top of the pole, with the bottoms of the cans against the pole. This made a good radar target, and in a calm sea the ship's radar would pick them up at a distance of several miles.

In using electronic methods to measure distance on hydrographic surveys it is necessary that the distances be measured from the point aboard ship where the transceivers for the echo-depth recorders are located. With Raydist, corrections are necessary. On the *Hydrographer*

the transmitting antennas for the 3280-kc/s signal was on the mainmast and the two receiving antennas were on the foremast. To correct the R_1 distance to the Red station to the foremast required the use of the correction formula

$$R_1 \text{ correction (in lanes)} = \\ 0.35 \cos (SH - A_r)$$

where SH is the heading of the ship by gyro compass and A_r is the azimuth from the ship to the Red station. The distance is measured from a point midway between the two masts, which are about 100 ft, or 0.7 Raydist lane, apart. To correct the R_2 distance to the Green station to the foremast required the use of the correction formula

$$R_2 \text{ correction (in lanes)} = \\ 0.35 \cos (SH - A_r) - 0.7 \cos (SH - A_g)$$

where the symbols have the same meaning as above and A_g is the azimuth from the ship to the Green station. The R_2 distance is measured by an elliptical formula wherein the sum distance R_1 plus R_2 is measured, and then the R_1 distance is subtracted in the phase meter. The R_1 distance in the sum distance is measured from the foremast; the R_1 distance subtracted in the phase meter is measured from a point midway

between the masts. The R_2 distance in the sum distance is measured from the mainmast.

To check the accuracy of the correction formulas the ship was anchored in Provincetown Harbor at the end of Cape Cod, and simultaneous sextant fixes and Raydist readings were taken on each 15° change in heading as the ship was swung about the anchor. With the ship's anglemen remaining in one position near the foremast, the correction determined should be in accordance with the above formulas. Figure 2 shows the correction curves, where the abscissa is the relative heading of the ship compared with the Red and the Green stations, and the ordinate is the correction in Raydist lanes. Both curves agree with the correction formulas.

A more interesting interpretation is obtained if the R_2 curve is plotted in polar coordinates as shown in Figure 3, where the radial lines are the supplemental angles of those used in Figure 2 and the distance from the center is the correction. A circle with a radius of 0.7 lane will coincide almost perfectly with the points denoting the corrections, indicating good agreement in the values. This point is especially significant when it is considered that the ship was more than 150 miles from the Green station. At the point where the test was made

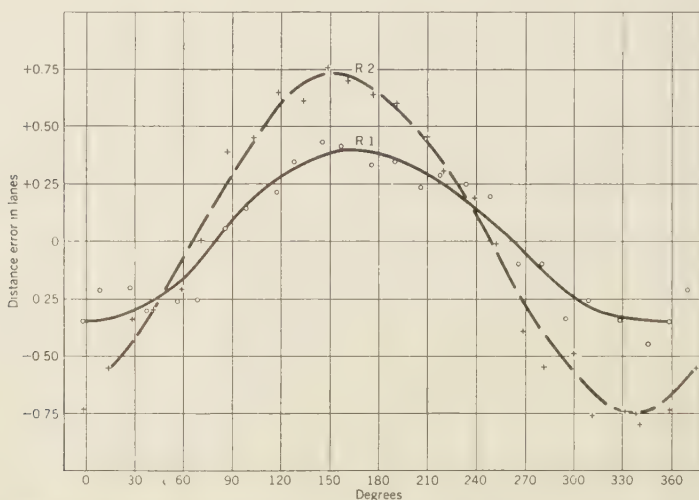


Fig. 2 Distance corrections as related to ship's heading. Tests made at Provincetown Harbor, June 4, 1958.

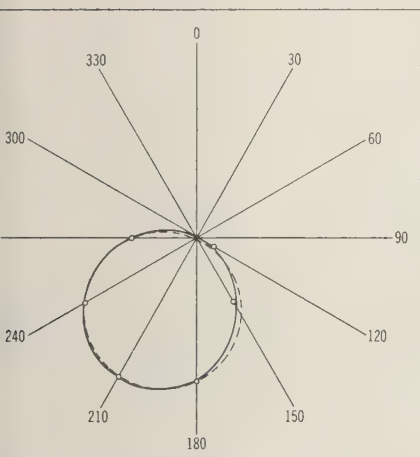


Fig. 3. R_2 corrections where R_1 station is 90° right of azimuth to R_2 station. Plotted azimuths R_2 station minus ship's heading.

station was almost exactly 90° to the right of the Green station. Under other conditions this figure would not be a circle, and the relationship of the figure would change.

A correction to the R_2 distance to the Green station may also be necessary, depending on the antenna arrangement at the Red station. In 1957 three 100-ft tower antennas were used, the 2496-kc/s transmitting antenna over the station center, the 1639.8-kc/s transmitting antenna 250 ft. distant on one side, and the 1639.8- and 3280-kc/s receiving antenna 250 ft. distant on the other side, the three antennas in a line. With this antenna arrangement there is a correction to the R_2 distance because the 1639.8-kc/s transmitting antenna is eccentric in the station center. The formula for the correction is

$$R_2 \text{ correction (in lanes)} = d \cos (A_s - A_i)$$

where d is the distance, in Raydist lanes, of the 1639.8-kc/s transmitting antenna from the station center (1.7 lanes in 1957), A_i is the azimuth

from the Red station to the ship, and A_s is the azimuth of the 1639.8-kc/s antenna from the station center.

In 1958 the antennas were rearranged at the Red station. A 35-ft whip antenna to receive the 1639.8- and the 3280-kc/s signals was placed on top of the 100-ft tower antenna used to transmit the 1639.8-kc/s signal, and the duplexed antenna was placed over the station center. The 2496-kc/s antenna was placed 250 ft off the center of the station, but the position of this antenna does not appreciably affect the distance. With the duplex antenna no correction is necessary to the R_2 distance as the azimuth of the ship from the Red station changes. The only objection to the duplexed antenna is that the strength of the 2496-kc/s signal has to be decreased to prevent interference with the duplexed antenna system. This limits the range at which usable signals can be received to about 125 to 150 miles.

In the DM Raydist as used on the *Hydrographer* the frequency of the transmitter on the ship determines the width of an R_1 lane, or half-wavelength.

For a frequency of 3280.000 kc/s the lane width is 149.87307 ft. The basic R_2 lane width is the same as for the R_1 , 149.87307 ft, but it is necessary to apply a correction as the ship moves away from the point where the dials were set to known values. The size of the correction is given by the formula

R_2 correction (in lanes)

$$= +0.00012[(\psi_R'' - \psi_R') - (\psi_G'' - \psi_G')]$$

where ψ indicates the phase meter dial readings, the subscripts R and G refer to the R_1 and the R_2 dials, respectively, and double prime and prime indicate the calibration point and the point for which the correction is desired, respectively. This correction is due to the elliptical method used to measure the R_2 distance and the fact that the frequencies are not exactly 3280 kc/s.

Some Developments in Loran

W. O. HENRY

*Electronics Engineering Division
U. S. Coast Guard Headquarters
Washington, D. C.*

Most of you are familiar with standard Loran (Loran-A) which has been in operation and providing radio-navigation service for nearly two decades. Briefly, Loran-A consists of transmissions of accurately timed, carefully shaped pulses of radio-frequency energy from synchronized master-slave stations to a user in the service area. The navigator's Loran-A receiver-indicator identifies the transmitting stations and measures the difference in microseconds between the times of receipt of the two pulses by comparison of their demodulated envelopes. The time difference thus measured locates the navigator on one of a family of hyperbolas whose foci are at the transmitting stations.

Loran-B and Loran-C are the developments in Loran to which the title of this paper refers. Each of these systems determines a coarse time difference in the same fashion as Loran-A but also provides a fine measurement of the phase of the radio frequency within the pulse.

Table 1 summarizes the characteristics of Loran-B and Loran-C. An examination of Loran-B and Loran-C separately will serve as a comparison of the two systems.

LORAN-B

The pulse characteristics, channel frequencies, and repetition rates used in Loran-B are exactly the same as in Loran-A. By adding phase measurement of the radio frequency within the pulse, Loran-B provides the high precision of a 2.0-Mc/s continuous-wave phase comparison system and yet retains the sky-wave and ambiguity-free features of Loran-A.

A typical Loran-B triad will appear as in Figure 1. Master-slave separation can be as much as 200 nautical miles with the present power capability of the transmitters. The single pulsed slave stations of the triad, S_m and S_s are synchronized with a bipulsed master, M , with all stations of the triad on the same recurrence rate so that a typical scope presentation

TABLE 1. Summary of Loran-B and Loran-C Characteristics

Characteristic	Loran-B	Loran-C
Frequencies, kc/s	1850, 1900, 1950	100
Bandwidth, kc/s	35	20
Pulse envelope	$A \cos^2 t, -\pi/2 < t < \pi/2$	$A t^2 e^{-2(t-1)}, t > 0$
Pulse rise time, μsec (10% to 90%)	21	46
Pulse repetition rate, basic	20, 25, 331/3	10, 12 1/2, 162/3, 200, 25, 331/3*
No. specific rates per basic rate	8	8†
Peak power, kw	200	300‡
Approximate ground-wave range, nautical miles, over sea water	200	1400
Standard deviation, μsec , repeatability	0.01	0.1
Signal-to-noise ratio required	10:1	1:10

*Group repetition rate. There are 8 pulses per group.

†Decrement of each specific repetition interval is 100 μsec .

‡Anticipated peak power for 75 kw at the sampling point.

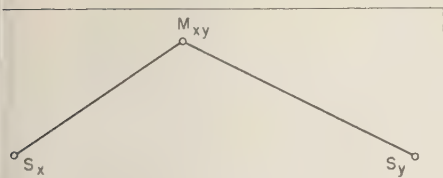


Fig. 1. Typical Loran-B triad layout.

on on a Loran-B receiver will appear as in figure 2. Time differences TD_x and TD_y are automatically and continuously measured by the receiver and presented to the navigator in read-outs of microseconds, or, with auxiliary equipment, as a cartesian plot or in meter read-outs of 'Left-right' and 'Distance to go.'

Loran-B receiver. The Loran signals from the antenna are fed via the coupling unit and transmission line to the TRF receiver in the equipment. This receiver has automatic gain control on a time-shared basis so that the g.c. level is automatically set for each master and slave signal. The amplified signals from the TRF receiver are then automatically applied to two servo mechanisms. The envelope time-difference servo measures the time interval between the reception of the master and slave pulse envelopes of each pair and displays these measurements as coarse indications. The phase-difference servo measures cyclic phase difference between the master and slave carriers of each pair and displays these measurements as fine indications. The two separate servos are interacting, however, and actually rack together.

The envelope and phase of the four signals,

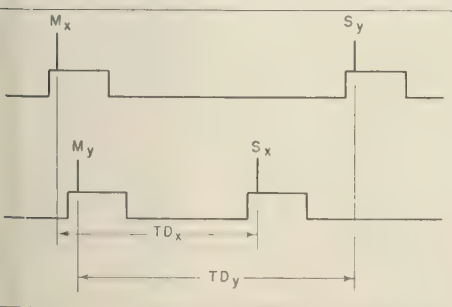


Fig. 2. Typical Loran-B scope presentation.

M_x, S_y, M_y, S_x , are applied to the envelope and phase servo systems on a time-shared basis, the time-sharing relays being programmed by the waveform generator. Only the operations required to obtain time difference TD_x will be described here, since TD_y measurements are obtained in a similar fashion.

The frequency divider generates the primary Loran cycle timing waveform used in the equipment, and the repetition rates of these timing waveforms are switched so that a complete program occurs during one repetition interval. The input to the frequency divider is the 100-kc/s signal from the oscillator. Zero reference for time-difference measurement is maintained by phase-shifting the 100 kc/s to the frequency divider to ensure that the zero reference and the M_x signal are time coincident.

The recovered envelope is applied to a differentiator circuit to obtain the first derivative of the Loran pulse. This first derivative of the M_x envelope is applied to the envelope error detector; the second input to the detector is a 4-microsecond strobe gate trigger from the waveform generator. If the strobe is not exactly centered about the zero crossover, the envelope error detector produces an error voltage which is amplified and applied to a servo motor in the time and phase difference measuring circuitry. The servo motor drives a resolver which phase-shifts the 100-kc/s input to the frequency divider, which, in turn, shifts the timing waveforms and strobe gate until the output of the error detector is zero. This condition obtains when the strobe gate is centered on the zero crossover of the first derivative of the M_x envelope.

With the zero reference thus established the equipment is ready to measure the $M_x - S_x$ time difference. An S_x strobe is delayed and applied to the envelope error detector coincident in time with the slope of the S_x derivative. The error voltage developed drives the S_x envelope servo until the strobe is centered on the zero crossover of the derivative. The time delay in microseconds between the M_x strobe and the S_x strobe is the coarse time difference displayed on the front panel drum counter.

The next operation to be performed on the Loran signal is the measurement phase. The output of the 100-kc/s oscillator is applied to

the frequency generator, where it is transformed to the carrier frequency. A 40-microsecond-wide burst of RF is compared with the phase of the master signal in a quadrature phase detector. An error signal will cause the phase servo to drive the resolver and phase-shift the 100 kc/s from the oscillator to phase quadrature with the M_s signal. At the same time an AFC capacitor is rotated in the oscillator. Thus the 100-kc/s oscillator output is always kept phase coherent with the master signal.

During the S_s sector an error voltage will be developed by the phase error detector if the S_s carrier and the frequency generator output do not have a 90° phase relationship. The S_s servo drives a resolver which phase-shifts the 100-kc/s input to the frequency generator to eliminate the error between the S_s carrier and the frequency generator output. Since the 100-kc/s input to the resolver is phase coherent with the master signals, the amount of resolver phase shift required to make the 100-kc/s signal coherent with the S_s signal is the amount of $M_s - S_s$ cycle phase difference. A drum counter is coupled to the resolver to display this phase difference in hundredths of a microsecond.

Loran-B ground stations. The function of the Loran-B transmitting stations is to transmit accurately timed pulses with the correct pulse shape and phase. The slave station timer compares the received local and remote pulses with locally generated timing and phasing references, and with appropriate servo systems maintains correct synchronization in time and phase with the master pulse. The master station transmits pulses of extreme stability with respect to time and phase and monitors the

round trip time on each slave base line. Techniques of timing, time sharing the signal paths, coherent detection, etc., described in the Loran-B receiver are similarly used in transmitting station timers.

The instrumental accuracy of the Loran-B equipment is such that time-difference measurements accurate to a few millimicroseconds are possible; the stability of the velocity of propagation, however, is not sufficient to yield a system accuracy to this precision. Measurements made to date show that the major variation in accuracy is attributable to changes in index of refraction, which is a function of temperature, pressure, and humidity of the earth's atmosphere.

The fact that Loran-B base lines are relatively short suggests that a single correction for velocity of propagation can be calculated using meteorological data for a given time, applied both to the triad and its service area. By siting the stations of the triad carefully as to eliminate propagation over inhomogeneous paths, or by further applying corrections to reduce the errors resulting therefrom, correction for velocity of propagation will produce a system repeatability of the same order as the instrumental accuracy.

LORAN-C

Loran-C operates in the frequency band from 90 to 110 kc/s with the shore stations on a triad or star chain transmitting on one of the repetition rates shown in Table 2 with pulse shape illustrated in Figure 3.

Figure 4 illustrates a typical Loran-C repetition interval, where T is one of the values shown in Table 2. In this arrangement, the slaves are synchronized with a single master

TABLE 2. Loran-C Repetition Intervals, microseconds

Basic/Specific	SS	SL	SH	S	L	H
0	100,000	80,000	60,000	50,000	40,000	30,000
1	99,900	89,900	59,900	49,900	39,900	29,900
2	99,800	79,800	59,800	49,800	39,800	29,800
3	99,700	79,700	59,700	49,700	37,700	29,700
4	99,600	79,600	59,600	49,600	39,600	29,600
5	99,500	79,500	59,500	49,500	39,500	29,500
6	99,400	79,400	59,400	49,400	39,400	29,400
7	99,300	79,300	59,300	49,300	39,300	29,300

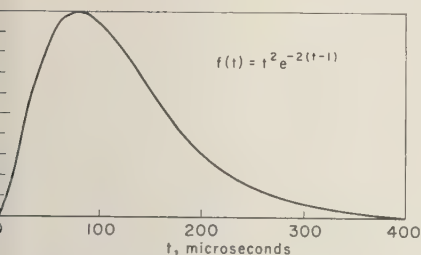


Fig. 3. Loran-C pulse envelope.

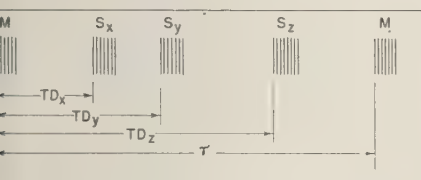


Fig. 4. A typical Loran-C repetition interval.

ion such as might occur in a star or square figuration.

In order to increase average transmitted power each station transmits 8 pulses spaced 100 microseconds apart at the group repetition rate. The pulses within each group are phase-shifted in accordance with a specified code as illustrated in Table 3. Phase coding greatly reduces the effects of sky-wave contamination, and, since the phase code varies from master to slave, it provides the necessary code for signal search, identification, and lock. It has been determined that at 100 kc/s the minimum ground-wave to sky-wave delay is about 30 microseconds. Sampling of a ground wave for envelope and phase must therefore be done before sky-wave contamination occurs at 100 microseconds, or approximately at the one-half amplitude point of the pulse.

Ground station operation. Loran-C equipment operation is described as follows:

(a) The master oscillator provides the stable basic frequency for station and network operation.

(b) The oscillator signal is passed through the phase coder and gates to provide transmitter excitation.

(c) The oscillator signal passes through the cycle compensator phase shifter to the divider. The cycle compensation phase shifter assures envelope and cycle coherence in radiated signals.

(d) The divider divides from 100 kc/s down to the basic and specific repetition rate for network operation, and furnishes triggers and gates for other basic timing functions.

(e) Gating circuits pass phase-coded reference signals to the quadrature (cycle) detector and through a 90° phase shifter to the in-phase (envelope) detector.

(f) The detected envelope is differentiated and added to itself to provide the derived envelope.

(g) Outputs from the in-phase and quadrature detectors are sorted by groups of time sharing relays.

(h) The appropriate sorted signals are integrated on the proper synchronous filter storage capacitor.

(i) The appropriate slave envelope or cycle servo is driven by the difference voltage appearing between the master and slave storage capacitors.

(j) The envelope servos drive the slave envelope time-difference units.

(k) The cycle servos drive the slave cycle phase shifters and indicators.

(l) The phase-coded 100 kc/s and triggers derived from the timer-synchronizer are passed to the transmitter and applied to the pulse generator.

(m) The pulse generator develops an envelope time function which, when acted upon

TABLE 3. Loran-C Phase Code

	Master								Slave								
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	
Interval I	+	+	-	-	+	-	+	-	+	+	+	+	+	-	-	+	+0° phase
Interval II	+	-	-	+	+	+	+	+	+	-	+	-	+	+	-	-	-180° phase

by the amplifier driver, the power amplifier, the transmission line, the antenna coupler, and the antenna, produces the desired radiated pulse shape.

In slave-station operation, the prescribed coding delay is manually set. The difference output between the master and appropriate slave envelope and cycle storage capacitors is utilized to drive the slave oscillator frequency control. The difference between slave envelope and cycle storage capacitors is used to drive the cycle compensating servo.

Additional circuits. Additional circuits are provided for automatic gain control, automatic gain limiting, semiautomatic search, sky-wave alarms, and off-synchronization alarms. Details of these circuits are not discussed here, since their operation is not an absolute requirement for station operation. A Loran-C triad is now in operation on the east coast of the United States with slave stations at Martha's Vineyard, Massachusetts, and Jupiter, Florida. The system was installed to provide a special-purpose navigation system for the U. S. Navy.

The transmitters and synchronizers installed at the ground stations were originally designed and built about 1946 for evaluation of a low-frequency Loran system utilizing pulse envelope matching techniques. The basic equipment was extensively modified to provide the more sophisticated instrumentation required for Loran-C. Extensive use of the breadboarded servo systems, wiring, etc., has been made since the system is only quasi-operational.

Receiver operation. The receiver circuit configuration is essentially the same as the ground-station synchronizer, except that oscillator frequency control servos are driven from the master envelope and cycle storage capacitors. Since, in a moving vehicle, there is a Doppler frequency shift, servos are designed to provide zero position error as long as velocity remains constant. Cross drive between cycle and envelope time-difference units is provided. Such cyclic-aided tracking allows the envelope measurement under quasi-static conditions.

System accuracy. Absolute accuracy in Loran-C is limited by several factors common to many radio navigational systems: (1) system predictability; (2) synchronization error; (3)

propagation anomalies; (4) receiver instrumental error.

If the Loran-C signals were propagated from precisely located antennas over the surface with uniform electrical characteristics, time differences for any point in the service area could be accurately determined and a receiver could then be accurately located at these points often as required. But, owing to the non-homogeneous nature of the transmission path from transmitting antennas to receiving antenna, there will be a secondary phase effect on the signals, resulting in errors in the time difference readings. These could be as large as 5 to 50 microseconds. Secondary phase factor corrections can be made and applied so as to reduce time-difference errors to about 0.5 microsecond. By calibrating the service area of the system and by precisely relating all points to the system to the same datum, prediction errors can be reduced to 0.05 microsecond. The errors mentioned above apply to Loran-A as well as to Loran-C. Secondary phase errors in Loran-C, however, tend to mask the other errors because of the longer propagation paths involved. Synchronization in normal system operation is maintained within 0.05 microsecond. Periodic variation of time difference readings due to instrumental and propagation anomalies is usually less than 0.1 microsecond.

Figure 5 shows standard deviation in time difference as a function of signal to noise ratio. With this curve and CCIR Report No. 34, Revision of Atmospheric Radio Noise Data, it is possible to construct contours of repeatability.

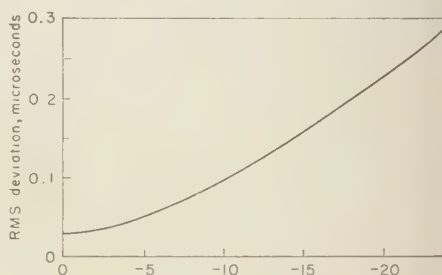


Fig. 5. Deviation of measured time difference due to white noise as observed on a typical Loran-C receiver.

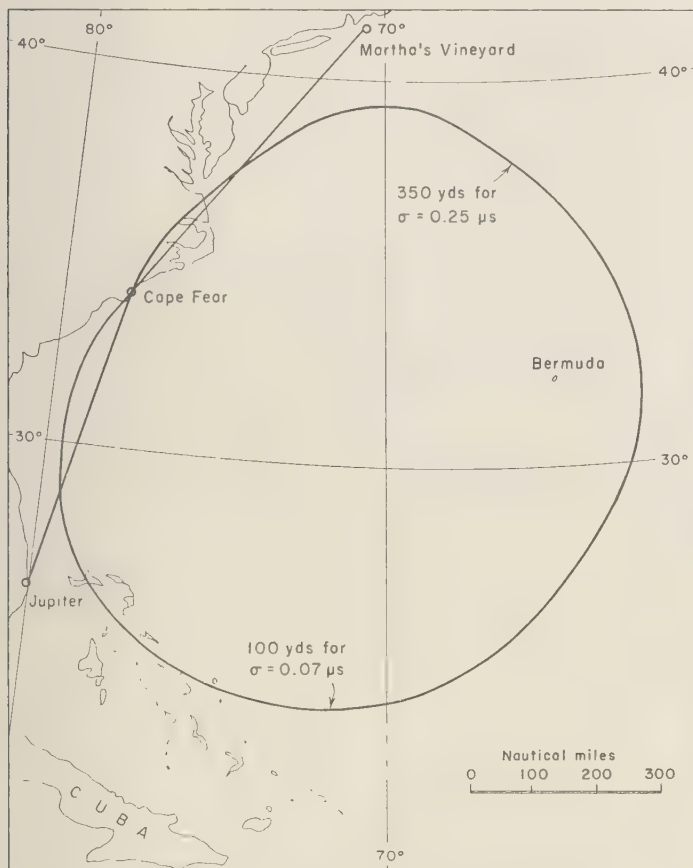


Fig. 6. East-Coast Loran-C.

TABLE 4. Approximate Ground-Wave Range 95 Per Cent of the Time for 100-Kw Peak Radiated Power

Geographic Area	Maximum Noise, db	User Equipment		Standard Deviation		
		0.1 μs	0.15 μs	0.20 μs	0.25 μs	0.3 μs
omer						
th Atlantic	35	1480	1600	1680	1780	1880
ter						
th Atlantic	40	1300	1460	1560	1660	1760
tral						
ific	45	1160	1280	1380	1480	1580
ibbean						
diterranean	50	1060	1150	1250	1350	1450
antic						
atorial	55	900	1060	1120	1220	1320

ity for any given area. Figure 6 is such a curve constructed for the east-coast Loran-C chain for the time block 0800-1600. On a year-round basis the total deviation due to noise alone in this time block is less than 0.07 microsecond if the radiated power level at the sampling point is 25 kw. Thus the east-coast Loran-C system corrected for velocity of propagation would produce the 95 per cent radial error contours of 100 yd illustrated. This is an illustration of the ideal. In practice, when deviations due to changes in propagation velocity as well as noise are considered, the 95 per cent radial distance error on the contour would be of the order of 350 yd, assuming that a service area calibration has been made.

Table 4 is a summary of ground-wave ranges within which a receiver would measure Loran-C time differences with the deviations indicated. These deviations are due to atmospheric noise alone and are therefore indications of system repeatability.

Up to this point only the ground-wave mode of Loran-C propagation has been considered. As would be expected from Loran-A experience, a considerable bonus in sky-wave coverage beyond the Loran-C ground-wave service area is available.

A thorough evaluation of the usefulness of Loran-C sky waves has been hampered because of a shortage of receivers, but a preliminary short-term investigation of sky-wave propagation using shipborne and airborne receivers indicates a stability that shows promise of application for navigation and possibly long-range distance measurement in hydrography. Emphasizing again that the results obtained are only tentative, they can be summarized as follows:

(a) The first-hop sky wave was useful to a distance of 2000 nautical miles during the day and 2300 nautical miles at night.

(b) Second-hop sky waves were tracked to 3000 miles at night.

(c) Third-hop sky waves were monitored solidly at night on the ground at 3435 miles, with every indication that the signals could be held for some distance beyond. Fourth-hop signals were also observed at this range and should carry well beyond.

TABLE 5. Summary of Loran-C Sky-Wave Stability

Mode	Predictability	Remarks
1st hop	$\pm 1 \mu\text{sec}$	4-hr time blocks excluding sunrise and sunset periods
2nd hop	$\pm 2 \mu\text{sec}$	4-hr time blocks night-time only

Note: The stability of 95 per cent of the readings based on a 4-hour mean in both modes is μsec .

(d) An almost complete darkness path between user and Loran-C stations is required for stable night-time multiple-hop reception.

(e) Sky-wave modes do not appear useful for navigational purposes during sunrise and sunset periods.

An estimate of the capability of using Loran-C sky waves is summarized in Table 5.

CONCLUSION

In conclusion, it can be said that Loran-C shows promise of being the most accurate, hour-a-day, over-the-horizon navigation system developed to date. It has the high precision of 2.0 Mc/s continuous-wave phase comparison system and yet will operate completely free of lane ambiguity or sky-wave self-contamination to a range of 300 or 400 nautical miles. Its present capabilities in existing equipment limit ranges to about 50 miles over land and 200 miles over water.

Beyond the Loran-B service area, Loran-C can provide the most accurate radio navigation service on a continuous basis that is known to be in existence today. Its ground-wave range has been given nominally as 1400 nautical miles for 100-kw peak power radiated; but the distance varies according to the atmospheric noise level of the area of interest and depends on the time of day and season of the year.

The fix accuracy at a point in the service area of either system depends, as with other hyperbolic systems, on the factor of geometric precision and the total deviation of the measurements differences observed at the point in ques-

possible to obtain the maximum accuracy either Loran-B or Loran-C for hydrographic surveying by applying corrections for changes in velocity of propagation and second-phase corrections to the observed time-dif-

ference readings. The burden for absolute accuracy, however, rests with the hydrographer, for he must orient precisely the hyperbolic grid of a Loran triad with respect to a given geodetic datum.

The Method and Use of Two-Range Decca¹

G. W. LACROIX AND D'A. H. CHARLES

*Surveys and Mapping Branch
Department of Mines and Technical Surveys
Ottawa, Canada*

1. INTRODUCTION

Two-range Decca is a variant of the Decca navigator system. It was developed specifically for surveying, and, unlike conventional Decca, which utilizes permanently based transmitters ashore, the master is carried in a ship and the shore-based equipment is designed for portability as well as for the essential accuracy. The method and use of Two-range Decca form the subject of this paper, but a few introductory remarks about the conventional system may be in order.

The Decca navigator system, now in use throughout Europe and in eastern Canada, has established itself as an aid to navigation of prime importance both to ships and aircraft. It can also be used effectively for surveying, and this application has been the subject of much work and several excellent publications.

In principle, a conventional Decca chain consists of a set of fixed-frequency radio signals generated by transmitters placed in strategic positions ashore. The loci of the points of constant frequency differences form hyperbolic patterns which are superimposed on normal navigation charts. Aboard ship, only receiving equipment is needed to provide the data from which accurate positions may be plotted without the necessity of computation. Lane identification is provided by the equipment, and any number of ships can use the chain simultaneously.

Four permanent Decca navigator chains are currently operated as aids to navigation in eastern Canada, and they have been successfully used by our ships for hydrographic surveys.

The accurate measurement of distance for position fixing by radio is a matter of keen interest to hydrographers. In 1954 the Canadian Hydrographic Service investigated the available equipment, and the following year a decision was made to adopt the Two-range Decca system for our work.

The first Two-range chain used in Canada was established aboard the *CGS Kapuskasin* in 1955, and work was started in the Gulf of St. Lawrence. This work was extended in 1956 and a second chain was set up in Cabot Strait using the *CGS Fort Frances* as the master ship. In 1957, our latest survey vessel, *CGS Baffin*, used the system off the coast of Nova Scotia and the *CGS Wm. J. Stewart* started Two-range Decca work in Hecate Strait, British Columbia. Both these projects were continued in 1958, and during that summer the *Baffin* transferred the Nova Scotia chain to the southwest coast of Baffin Island, where it was used to survey previously uncharted region.

In the following, the method of Two-range Decca is briefly outlined and its use as an aid to offshore hydrography is discussed. The calibration trials and other tests described are based on the experience gained during two seasons of work on the Pacific coast, for which the equipment provided was a Two-range system of the 6f, 300-watt type, using 'Green' and 'Purple' slaves.

2. THE METHOD OF TWO-RANGE DECCA

In a Two-range Decca chain, the master transmitter and a Decca receiver of conventional type are carried in the ship whose position, relative to two slaves ashore, is to be established. The decimeters, associated with the receiver, indicate the phase difference between the master signal and that received from

¹ Presented by permission of the Director, Surveys and Mapping Branch, Department of Mines and Technical Surveys, Ottawa, Canada.

slave. This phase reading bears a linear relation to the length of the transmission path and hence is a measure of the distance. The position lines are circles concentric about the ship, and only the ship carrying the master station can utilize the chain.

Considering a Two-range chain, with Purple and Green slaves of the 6f type, the basic frequency may be called f . The master aboard ship radiates frequency $6f$, and this is received by the Green slave, whose drive is phase locked at $6f$ and then radiated from Green at frequency $18f$. In the ship's Decca receiver, this received frequency is raised to $18f$ and compared with the master frequency, also multiplied to $18f$. The phase comparison is accomplished within the phase discriminators, and the phase difference is displayed on the 'Green' decometer.

In like manner, the Purple slave receives frequency $6f$ from the master and returns it at frequency $5f$. This frequency is raised to $30f$ and phase-compared with the master frequency, and the phase difference appears on the Purple decometer.

The Decca measurements of any chain are governed by the assigned frequencies and the speed of wave propagation. The half-wave-length, called lane width, is half the wave velocity divided by the comparison frequency.

The velocity of radio waves at Decca frequencies over sea water, as presently accepted for our purpose, is 299,650 km/s. However, since they travel as surface waves, their speed is influenced by the electrical properties of the surface medium. Conditions over sea water are fairly uniform, but not over land. In general, the speed is less over land than over sea water.

For hydrographic work it is evident that slave stations should be located so that, as far as possible, a wave path over land is avoided. The Decca Company recommends that the slaves should be located within 300 meters of the shore whenever possible.

Since the comparison frequency of Green is $18f$ and that of Purple is $30f$, it follows that 18 Green lanes are equal to 30 Purple lanes, and this is called a zone. The Green decometer is numbered in lanes from 30 to 48, and the zones are indicated by the letters A to J. Thus the possibility of confusing zone A with zone A plus 10 is remote. The Purple decometer is

numbered 50 to 80 lanes and the zones are also lettered A to J. In this way, confusion between Green and Purple decometer readings is avoided.

No lane identification is provided in our Two-range systems. At the start of work, the proper number of whole lanes must be set on the decometers at a point where the Decca coordinates are known. After that, the decometers keep the lane count automatically.

For practical surveying, the following approximate equation may be used:

$$R = D + a + b$$

where

R = distance (lanes) between effective electrical centers of master and slave.

D = decometer reading (lanes) to which the mean goniometer correction has been applied.

a = locking constant (lanes).

b = correction due to proximity of slave's induction field (lanes), from Norton and Bremner curves.

The small error which may occur as the result of goniometer change with variation of decometer readings can be neglected. It may also be assumed that the wave speed remains constant if care is taken to ensure that the wave path is largely over sea water and free from significant electrical interference.

The effective electrical center at slaves may be taken as the mid-point between transmitting and receiving aerials if there is no significant electrical interference and the separation of aerials is about 200 ft.

At the master aboard ship the aerial separation will often be much less than 200 ft, and the effective electrical center cannot be closely defined as the mid-point between them. It can be found by trial, however, as will appear in section 4. Its position may also vary a small amount about a central point, depending on the bearing of the ship relative to the slave, but this effect is small and need not be considered.

The locking constant is a significant correction introduced by the ship's induction field. Its magnitude may vary but, for a given set of conditions, can be considered constant for practical purposes. This effect increases the decom-

eter readings. It must be determined by trial (section 4) and then subtracted from the decometer readings by offsetting the goniometers.

The correction due to the slave's induction field is a function of distance. If the master aboard ship approaches within about 6 miles (11 km) of the slave, it must be considered. Like the locking constant, its sign is negative; that is, the amount must be subtracted from decometer readings. This correction has been computed from theory and is determined by the use of Norton and Bremner curves (Fig. 1).

The best possible positions of the slave electrical centers must be obtained by surveying, as all subsequent Decca measurements depend upon them.

The nondirectional, unmodulated, continuous-wave signals used in Decca chains are in the frequency band 70 to 130 kc/s. At these frequencies the waves tend to follow the earth's surface, and so the range of a Decca system is not limited to the optical horizon.

The Decca frequencies assigned for use on the Pacific coast and the resulting lane widths and comparison frequencies are:

Master, kc/s	85.046
Purple slave, kc/s	70.872
Green slave, kc/s	127.569
Purple comparison frequency, kc/s	425.230
Green comparison frequency, kc/s	255.138
Purple lane, meters	352.3387
Green lane, meters	587.2312

A list of the equipment required for the 6f, 300-watt Decca system is given in the Appendix.

For plotting the Decca results, a special method was developed by the Nautical Geodesy Section of the Hydrographic Service.

3. SUMMARY OF OPERATIONS

The assignment was in Hecate Strait, where an area exceeding 6000 square miles, within Capes Scott and St. James, had not been adequately surveyed.

The shores of this strait are rocky and broken, and the soft-wood forest, almost without exception, grows to the water's edge. This, together with the tangle of fallen trees and underbrush, caused the siting of the slaves to be rather a severe problem. It was only overcome after several weeks' work at each place,

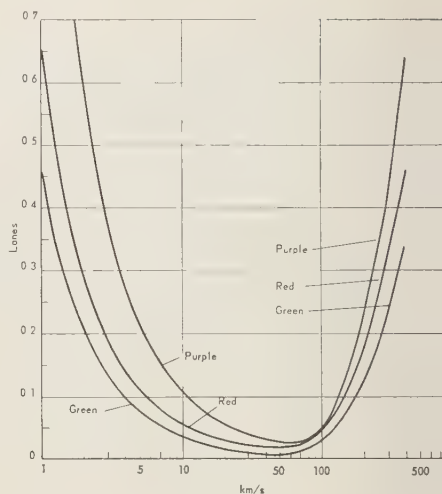


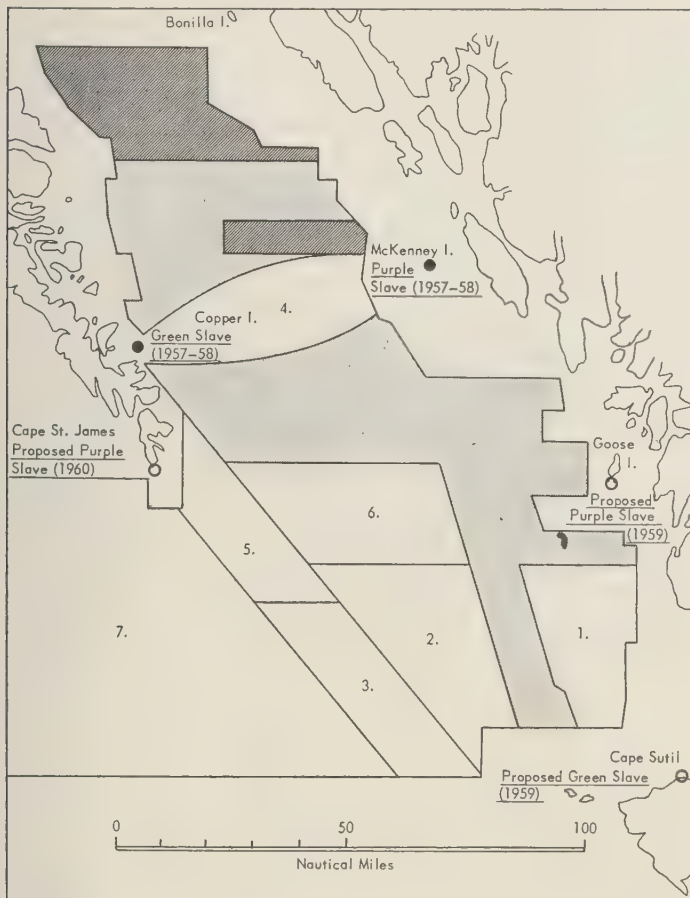
Fig. 1. Phase correction over a sea-water transmission path (conductivity $\sigma = 5 \times 10^{-11}$ em

and the locations are by no means ideal, but they were the best available.

The Purple slave was placed on a small island in the region west of Aristazabal Island and the Green slave on Copper Island, at the entrance to Skincuttle Inlet, Queen Charlotte Islands (Fig. 2). The locations provided wave paths over the sea without intervening land masses of any consequence, and these slaves were used for all the subsequent work in 1954 and 1958. Prefabricated houses were used for the slave equipment and personnel, and the generators were placed under cover in separate shacks.

When the slaves were ready for operation the first step was to determine the geographical positions of their electrical centers. This was done by surveying from previously established triangulation points ashore. From trial observations, the ship's electrical center was found and the locking constants were determined. The decometers were corrected accordingly, and the system was then, presumably, ready for use. However, it appeared essential to make sure that the new equipment would fully satisfy the standard of accuracy required, and for this reason additional trials were made before surveying was commenced.

First, the distance between the slaves as



LEGEND

- Present slaves
- Proposed slaves
- Surveyed 1957
- Surveyed 1958
- 1 to 7 Future work

SEQUENCE

- April 1959—Green at Copper I. Purple at Goose I. Complete area 1.
- June 1959—Green to Cape Sutil. Complete areas 2 and 3.
- September 1959—Purple to McKenney I. Complete areas 4, 5 and 6.
- 1960—Green at Cape Sutil. Purple at Cape St. James. Complete area 7.

Fig. 2. Area surveyed 1957-1958 and proposed future work.

ed from the geographical positions was compared with the distance as determined by Decca. When a final test was conducted near the northern limit of the proposed work where a position determined from a shore base was compared with the Decca result. The close coincidence obtained by these tests indicated that very good results could be expected from the Decca sys-

tem, and it was then considered safe to proceed with the work. The details of the above trials are given in section 4, and the results, together with other trials and computations made in 1958, are discussed in section 5.

The correct lane numbers were set into the decometers either by visual fixing, from range marks to natural objects ashore, or by the use

of buoys for which the Decca coordinates had been previously established. Offshore, the buoys proved to be essential. They were moved to strategic locations from time to time as the work progressed.

CGS Wm. J. Stewart, the survey vessel used for the work, was built in 1932. Her length is 228 ft, breadth 35 ft, mean draft 13½ ft, gross tonnage 1295, and normal cruising speed 10 knots.

As a result of the two seasons' work, 10,252 nautical miles were sounded in Hecate Strait using the Decca Two-range system. The area now completed, as well as that still remaining to be done, is shown in Figure 2.

4. BASE POSITIONS AND DECCA CALIBRATION

Base positions. The position of Green electrical center was obtained from a near-by geodetic base through two independent nets. Four triangles at most were involved, and the results checked closely. It is considered that the accepted position is of a high order, probably coinciding with a primary geodetic value within 2 ft.

The position of Purple electrical center is less certain. It was derived from a base established by the Hydrographic Service in 1922. Two values were obtained by different routes, which showed a negligible difference, and it was evident that the accepted position is just as valid as the survey of 1922 on which the charts are based.

It would have been preferable to start from the geodetic base, but this could not be managed without unduly delaying operations. The side used was only six triangles distant from the geodetic base of the original survey, however, and the position error of Purple is unlikely to be great. Nevertheless, it is recognized that residual error probably does exist in this third-order triangulation.

Ship's electrical center. As no wharf was available, a location near Purple was chosen where the ship could safely steam toward the slave, and in a reverse direction. The line of sight of a theodolite placed ashore at right angles to this course provided the fix. Radio telephone communications were maintained from ship to shore, and the decometers were read at the moment the observer indicated the

TABLE 1. Ship Electrical Center Observations; Purple Decometers

Ship's Head, gyro	'Standard'	'Reference'	Half Differences, lanes	Equivalent Feet
322	0.267	...	0.0435	50
142	0.354	...		
322	0.255	0.288	0.0575	66
142	0.373	0.400		
322	0.300	0.339	0.0325	37
142	0.370	0.400		

Mean distance, foremast to ship's E/C, 0.044 Purple lane, 51 ft. This value was used in all subsequent computations.

foremast to be on his line of sight. A slow ship's speed was maintained, and calm weather prevailed.

The results, listed in Table 1, did not agree too well. The mean value is likely to be quite close to the true value, however, and it was accepted.

Locking constants. To determine the locking constants, 'ship stations' were conducted in which observers positioned the ship's foremast from a shore base with theodolites. The observations were synchronized by dipping a flag aboard ship at prearranged intervals and simultaneously reading the decometers and the ship's bearing. Ten repetitions were made in each case and the results averaged. The locking constants were then computed from the approximate equation given in section 2.

$$R = D + a + b$$

Since R is the distance between the electrical centers of master and slave whereas the distance resulting from the shore-based observations and subsequent computations is that between the slave's electrical center and the ship's foremast, the latter was corrected as follows

$$R = R_1 + X \cos \theta$$

where

R_1 = surveyed distance from slave electrical center to foremast.

X = distance, previously determined, foremast to ship's electrical center.

θ = angle between the bearing of ship's head (observed) and the bearing of the slave from the ship (computed).

Thus,

$$R_1 + X \cos \theta = D + a + b$$

which $X \cos \theta$ takes the sign of the angle, and a and b are always negative. The b correction was taken from Norton and Bremner curves (Fig. 1), and the locking constant, a , could then be computed.

Example, Locking constant computation

R_1 12.812 Green lanes	12.812
D 13.286 Green lanes	0.003
b 0.047 Green lanes	0.047
X 0.026 Green lanes	
θ 84°	12.862
	-13.286

a The 'Green' locking constant -0.424

Locking Constant Results, 1957, 'Standard' Decimeters

Trial	Purple	Green
1	-0.763	-0.424
2	-0.794	-0.443
3	-0.556	-0.409
4	-0.600	-0.545
5	-0.592	-0.439
Means accepted	-0.661	-0.452

Locking Constant Results, 1958

Trial	'Standard' Decimeters		'Reference' Decimeters	
	Purple Lanes	Green Lanes	Purple Lanes	Green Lanes
1	-0.577	-0.407	-0.578	-0.421
2	-0.602	-0.495	-0.654	-0.514
3	-0.572	-0.377	-0.574	-0.387
4	-0.588	-0.372	-0.612	-0.389
5	-0.582	-0.506	-0.592	-0.524
6	-0.626	-0.436	-0.629	-0.443
7	-0.684	-0.387	-0.704	-0.402
Means accepted	-0.604	-0.426	-0.620	-0.440

Base-line distances. The distance between the slave electrical centers, as computed from

surveys by the inverse method, was 64.42 nautical miles. This is equivalent to 339.107 Purple lanes.

To test the total lane count, a trial run was made from Green to Purple, starting and ending with sextant fixes. This showed the integral lane count of both decimeters to be in accord with the above measurement.

A further base-line measurement was made as follows: Starting near Purple, lane identification was accomplished and the ship's course made good on the base line toward Green. When beyond the influence of the slave's induction field, the total Decca distance, Green plus Purple, was recorded at quarter-hour intervals over a 2-hour period. The ship would not have wandered far from the base line during this time, and the observations were discontinued before the onset of the sky wave. The mean base-line value thus obtained was 339.117 Purple lanes, which is 12 ft greater than the computed value. It is to be noted, however, that the 1957 locking constants were used in this trial.

Further Decca measurements of the base line were made in 1958, the results of which appear in Figure 3, and this matter is discussed in section 5.

Final position check. This trial, made in 1957, was considered to be the most valuable check on the performance of the new equipment. The place chosen for the test was approximately 60 nautical miles from Green and 90 from Purple. It was thought that if a good check could be obtained here it would, in large measure, confirm the reliability of the Decca system for the work in hand.

A 'ship station,' consisting of the usual ten repetitions, was conducted, and the 1957 locking constants were used to determine the Decca distances.

As before, a geographical position of the ship's foremast was derived from a triangulation base ashore, and the position of the ship's electrical center was computed.

Using the Decca values, another position of this electrical center was found as follows: Given the side Green to Purple as computed by inverse and the observed Decca distances from the ship to the slaves, the angles of this large triangle were found and the ship position solved using the Green and Purple positions

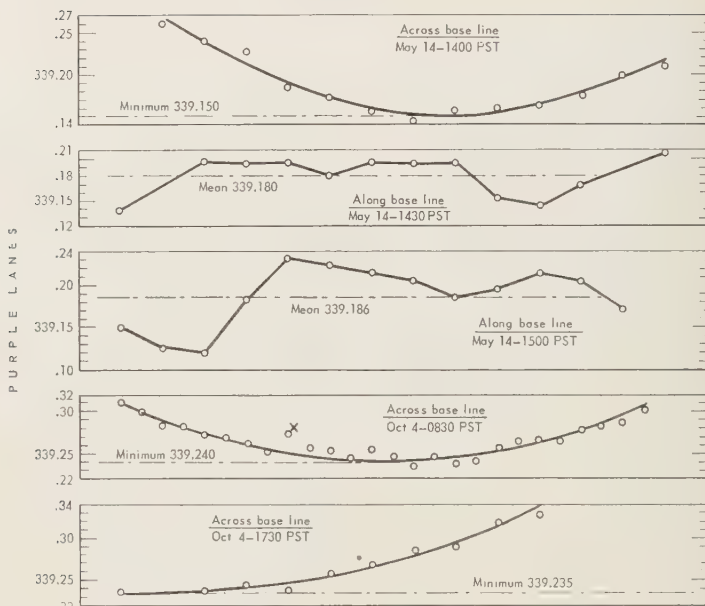


Fig. 3. Base-line tests 1958.

which had been accepted from the surveys.

The resulting difference was 82 ft (25 meters), which is equivalent to 0.013 in. on the 1/72,688 scale plotting sheets. This trial is further discussed in section 5.

5. ADDITIONAL OBSERVATIONS, COMPUTATIONS, AND CONCLUSIONS

The 1957 locking constant results were none too consistent, and the trials were repeated in 1958 (section 4). With the previous experience in mind, every precaution was taken to try to obtain more uniform results. The number of trials was increased; all observers were requested to be as precise as possible; and both the 'Standard' and the 'Reference' decimeters were read in order to establish locking constants for each pair. The 1958 results were better than those of 1957 and are considered satisfactory for all practical purposes. Quite a spread still appeared in the individual results, however, and the reason for it is obscure.

The differences between the 1957 and 1958 locking constants amounted to 0.057 Purple lane (66 ft) and 0.026 Green lane (50 ft). If it is assumed that the 1958 values are correct

for both years, the 1957 Decca results would be in error by the above amounts. No consideration was given to replotting this work as any changes would scarcely exceed 0.015 on the plotting scale that was used.

The results serve to indicate the necessity of careful evaluation of the locking constants because any error introduced into the system during calibration will be carried into all work which follows.

Recomputation of 1957 position check. Before stated, the 1958 locking constants were believed to be better than those of 1957. As it appeared reasonable to suppose that the true values could not have changed materially since the same slaves were used in both years and no alterations were made to the equipment. It was reasoned that insertion of the 1958 locking constants into the Decca values of the position check made in 1957 (section 4) should result in a closer coincidence between the geographical coordinates obtained by Decca and by the survey. This reasoning proved to be correct. When recomputed, the difference in position was reduced from 82 ft (25 meters) to 18 ft (5½ meters).

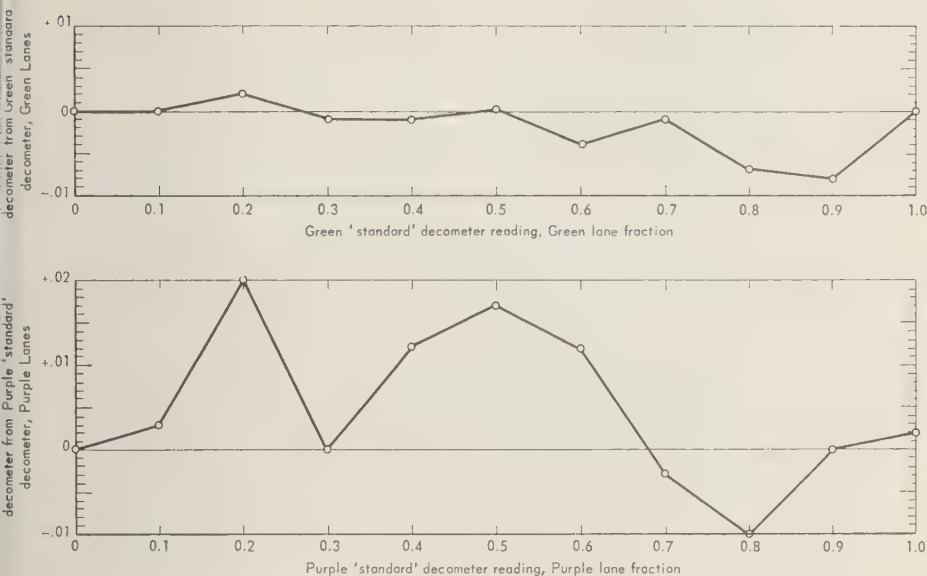


Fig. 4. 'Nonlinearity' of Decca receivers.

Base line. Referring to the 1958 Decca measurements of the base line (Fig. 3), it was evident that these results were not very consistent, varying from 339.150 to 339.240 Purple lanes. It was estimated that the May 14 trial was probably the best, and a Decca base-line value of 339.160 Purple lanes was used for the computation of wave velocity.

As a result of the base-line trials, it seems evident that the best way to measure a Decca base is to steam the ship slowly across it, reading the decometers at short intervals. A smooth curve drawn through the plotted points then serves to remove the scatter inherent in the individual Decca readings, and the minimum value is, of course, the desired measurement. This method will be used in future.

Variation of decometer readings. A difference in the locking constants derived for the two Green decometers, and also for the Purple pair, was noted in the results (section 4), and to investigate this a further trial was made. By means of a signal generator, the same signal was impressed on both Decca receivers and the decometer differences were noted at various points around the dial. This procedure showed that the Decca receivers are not per-

fectly 'Linear'; that is, the decometers do not read exactly the same at all times. The Decca Company states in its manual that this difference may amount to as much as 0.02 lane, a value largely confirmed by the above results as shown in Figure 4.

It is evident that small errors, of the order of 0.02 lane, are inherent in the equipment and may be expected to occur at any time.

The variation of individual Decca readings from all sources was demonstrated by the trial made along the base line in 1957 (section 4), where the extreme variation of nine readings from their mean was ± 85 ft (0.074 Purple lane). The pattern of such variations is also illustrated in Figure 3, where the readings, made at 30-second intervals crossing the base line, are plotted.

Repeatability of Decca readings. The appropriate tests having been made, and the work started, it was desired to know just how well the new system would remain in adjustment under operating conditions that varied a great deal from time to time. A conclusion was not immediately available, but within a short time an impressive array of evidence was gathered from the daily position checks. Some typical

TABLE 2. Differences in Repeated Readings at Check Points; Buoys Anchored in Depths up to 25 Fathoms

Purple Lanes	Green Lanes	Purple Lanes	Green Lanes
0.03	0.03	0.13	0.09
.16	.10	.08	0
.06	0	.08	.07
.11	.10	.18	.01
.11	.02	.01	.02
.13	.01	.06	.08
.06	.03	.09	.06

examples are listed in Table 2. They are the differences obtained when the ship checked back at marker buoys placed in relatively shallow water for which the Decca coordinates had been previously established and fairly represent the degree of coincidence that was almost always obtained.

The excellent repeatability of the Decca system was also demonstrated many times by the ease with which the ship could return to the buoys. On one occasion when it was nearly dark the ship almost ran down the buoy before it was sighted from the bridge.

Extreme range. To test the Decca result at an extreme range, a 'ship station' was conducted in 1958 near Cape Sutil (Fig. 2) at a mean distance of 125 miles from the slaves, with an intersection angle of 27° . The result showed a difference between the Decca and surveyed positions of 367 ft, or 0.06 in. on the 1/72,688 scale plotting sheet.

At the extreme southern limit of soundings (Fig. 2), where the mean distance from the slaves was around 105 miles and the intersection angles varied from 31° to 35° , it is fair to assume that this difference would be somewhat less, probably about 0.05 in. as plotted, according to the ratio of distances.

During this ship station the Purple signal was noted to be very weak, and this fact, together with the magnification of residual error in the surveys, may account for the rather disappointing result. The test served to indicate that extreme distances and poor intersection angles are to be avoided whenever possible.

With regard to the sounding that had been done at the southern end of the area, it was decided that the probable positioning error

could be tolerated under the circumstances since all the soundings indicated a very flat bottom with depths of 30 fathoms or more throughout.

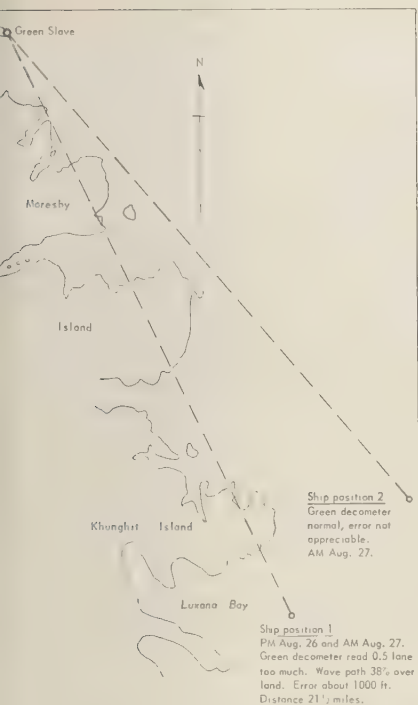
Effect of land masses on the Decca wave speed. This effect was so clearly shown on one occasion that it is of interest to record the event. On August 26, 1958, at the close of the day's work it was convenient to anchor for the night at 'ship position 1' (Fig. 5). Visual fixes showed that the Green decometer was reading too high by almost exactly half a lane. The next morning, in the same position, the fixes again showed this discrepancy. The decometer was set as before, however, on the assumption that the increase in speed of the wave caused by the intervening land had increased the decometer reading. This proved to be so. When the ship reached 'Ship position 2,' well clear of the land, additional fixes showed clearly that the Green decometer had returned to normal, and checks were obtained.

As shown in Figure 5, the Green wave speed was about 38 per cent over land, and this land is mountainous and believed to be rather highly mineralized. It is of interest to note that the error introduced to the Decca reading on this occasion was nearly 1000 ft, and it is doubtful that sky-wave effect could have accounted for more than a small fraction of it.

Velocity of radio waves at Decca frequencies. Throughout the work, both in 1957 and 1958, a Decca wave speed 299,650 km/s was used, and the position circle templates and plotting diagrams are made up on this basis.

Several different values of Decca wave velocity have been used by different authorities, and it was of interest to attempt evaluation of the wave speed from the 1957 and 1958 data. The distances used were those determined from the surveys. The computations, some of which involved very long sides, were well checked and are substantially correct. The 1958 locking constants were applied to the Decca readings in all cases.

The values were computed on the following basis. Assuming that there are no survey errors and that the Decca constants of 1958 are correct for both years, the total distance measurement by Decca is half the wave speed divided by the comparison frequency multiplied by the total lanes as read on the decometer. The



5. The effect of a land mass on the Decca wave velocity.

ould equal the distance obtained by the sur-
The other values of the equation being
wn, the wave speed was solved.

Result of Computations

Side	Length, km	Wave Velocity, km/s
e line	119.2	299,602
ple to ship 6, 1957q	163.0	299,660
en to ship 6, 1957	105.1	299,660
en to ship 9, 1958	259.4	299,531
ple to ship 9, 1958	217.4	299,496

it is not clear why the values determined
n ship 9 should fall so low. Possibly they
due to magnification of small survey errors
the very long distances and the small inter-
tion angle or to phase changes occurring
r the Decca wave path. At any rate it ap-
ers evident that they are not to be trusted.
uncertainty exists in the other three
es, as the Decca signals were strong, the
cca base measurement is at least close, and
p 6 was in a well conditioned triangle so

that residual survey errors could not have been greatly magnified. They should agree fairly well with the true Decca wave speed, and it is noted that their average is 299,641 km/s as compared with 299,650 km/s which is the value being used.

Adequacy of the Two-range Decca system. Although much of the Decca equipment is duplicated, a few breakdowns occurred from time to time during the Hecate Strait survey. They were of the type common, no doubt, to all complex electronic equipment, such as tube failures, blown fuses, failure of other components, and the result of moisture collecting on aerial tuning coil. The technicians always repaired the trouble without delay, and they deserve much credit for the success of the operation. Only a very small part of the time was lost because of Decca operational faults.

The lack of any means of lane identification within the system was inconvenient. Each day, after identification was accomplished, a constant watch had to be kept on the decometers to make sure that no lanes were lost, particularly when atmospheric conditions created a poor signal-to-noise ratio. However, the lane count was in error only twice in Hecate Strait when the ship arrived back at a check point in the evening, and very little of the work had to be discarded. On a few days, faults developed and lane integration was lost when the ship was many miles offshore. Then it was necessary to return to the nearest check point, and time was lost.

The chief problem experienced with the system used aboard the *Wm. J. Stewart* was the relative inadequacy of the Purple frequency as compared with the Green. Although it usually functioned well up to 70 or 80 miles from the slave, the Purple signal invariably became weak at greater distances. On a few occasions it was totally unreliable, and the work was interrupted. When atmospheric conditions were poor, Purple occasionally weakened even at distances less than 70 miles, causing a disconcerting oscillation of the decometer needle and even skipping of lanes.

At no time, nor at any distance worked, was trouble of this nature experienced with the Green frequency. Both slaves use exactly similar equipment, and it is certain that, under these conditions, Purple is less reliable than

Green. It was concluded that the range from Purple slave should be restricted to about 75 miles until some way is found to improve the signal strength. Possibly this can be done by raising the slave transmitter aerial.

The most prevalent difficulties encountered during the east-coast surveys include thunderstorms, interference from other radio transmissions, weak signals, and grounding of the Decca mast on board ship.

It was found that radio interference could usually be overcome by lengthening the slave transmitter mast or by increasing the top dressing of the master transmitting aerial, thus increasing signal strength. On certain occasions, interference was traced to reradiation from other shipborne aerials and overcome by relocation of the offenders.

Grounding of the master transmitter mast was overcome by fitting cowlings of copper or Fiberglas to shield the base insulator from salt-water spray, heavy rain, and snow.

Exceptional, and as yet unsolved, interference from radio transmissions last year on the *Baffin's* northern survey caused a change from Purple-Green to Red-Green slaves for the future operations.

The Two-range system is not affected seriously by fog, rain, snow, or rough weather, and, on the North Pacific coast, electrical storms are rare. It has functioned quite well even in weather so bad that the echo sounders failed to deliver depth readings, and, despite the imperfections mentioned above, it is fair to say that the equipment remained in satisfactory working order about 95 per cent of the time.

The need for constant vigilance and daily checks when using this electronic system is self-evident. The value of careful calibration before use and the danger of overextending the range have been indicated. With due allowance for these matters, it is concluded that Two-range Decca is a powerful aid for offshore hydrography.

The Two-range chains operated by the Canadian Hydrographic Service have proved dependable to ranges beyond 100 miles. When using the Decca Navigator chains of Nova Scotia and Newfoundland, the range could be extended out to 250 miles, and one ship completed 17,000 miles of sounding in a season. During the 4 years the Decca systems have been

used for Canadian surveys, almost 40,000 miles of sounding have been accomplished. Great savings of time have been effected, and ranges never before possible have been employed.

APPENDIX

List of equipment, Decca Two-Range System 6f, 300-Watt Type

Master Equipment, Shipborne

Transmitter aerial, 60 ft, lattice type, light alloy guyed mast, top capacity loading, vertical radiator, on base insulator, fed from tuning network.
Aerial loading coil, on horizontal stand.
Master control unit, double master type 793.
Transmitter unit, 300-watt, type 388.
Power unit, 300-watt, type 2904.
Converter, 110 volt DC to 230 volt AC, single phase, 60-cycle.
Two receiver units, type 375.
Three decometer units, survey type.
Two receiver aerials, plain insulated vertical wires, 40 ft long.
Track plotter, type 350.
Control amplifier for track plotter, type 388.
R/T communication set for ship-to-slave communications.

Spares and Accessories

Spare power amplifier unit.
Spare drive unit.
Signal generator, type 306.
Valve voltmeter.
AC-DC meter 'Avo.'

Note. The whole unit operates on 230 volt AC, and the over-all power consumption is 150 watts.

Equipment for Each of Two Slaves

Transmitter aerial, 100 ft tubular, light alloy guyed vertical radiator in 10-ft sections, top capacity loaded.
Aerial loading coil, horizontal stand.
Earth mat, one hundred 100-ft copper wires attached to base of mast forming radial earthing mat.
Two generators, each 3-kva output, diesel Newage-Alcon.
Two slave control units, type 793.
Transmitter, 300-watt, type 388.
Power unit, 300-watt, type 2904.
AC distribution box.
Receiver earth mat, 12 to 20 radial wires, copper, 50 ft long.
Two receiver aerials, whip type, insulated, 50 ft long.
RT communication set for slave-to-ship and slave-to-slave communications.

Note. Each slave has the same spares and test gear as the master.

Transistorized Raydist as Used in Geological Surveys¹

CHARLES L. DRAKE AND WALTER C. BECKMANN

*Lamont Geological Observatory
Columbia University
Palisades, New York*

Introduction. During January 1958 a survey was made across the southern reaches of Chesapeake Bay as a part of the foundation studies for a proposed highway crossing. Included were a sounding program and a survey to determine the nature and configuration of the sedimentary layers by sonic methods. The borings were made by Raymond-Tidewater, and the sonic apparatus was provided and operated by Alpine Geophysical Associates, Norwood, New Jersey. Navigation and position were controlled by a raydist, Hampton, Virginia. Over-all responsibility for the work rested with Sverdrup and Corcoran, St. Louis, Missouri, Consulting Engineers to the Chesapeake Bay Ferry District, which sponsored the project.

The problem. The preliminary investigations of foundation conditions along the line of the proposed highway crossing called for rather widely spaced boreholes. Since the boring sites were spread over a distance of 20 miles and located in the open waters of Lower Chesapeake Bay, efforts were made, in the interest of economy, to minimize the total number of holes to be bored. Investigations made in connection with other projects in the upper reaches of Chesapeake Bay suggested the possibility that deep channeling might exist in places along the bottom of the proposed crossing. The purpose of the sonic survey was to fill in the spaces between the boreholes so that no such anomalous structure could cause difficulties during the subsequent construction and to outline critical areas that required additional borings for further analysis.

Since the highway crossing was to follow a definite track it was important that the borings

be located as close as possible to this track. Furthermore, since interpretation of the sonic data required correlation with the borings, it was especially important that the relative positions of sonic data and boreholes be ascertained. Once the borings were completed and the drilling platforms withdrawn, the positions of the holes were unmarked and could be located only by precise navigation.

The area. The entrance to Chesapeake Bay is marked by strong and variable tidal currents with a wide exposed mouth, open to the winds and waves of the Atlantic Ocean (Fig. 1). Al-

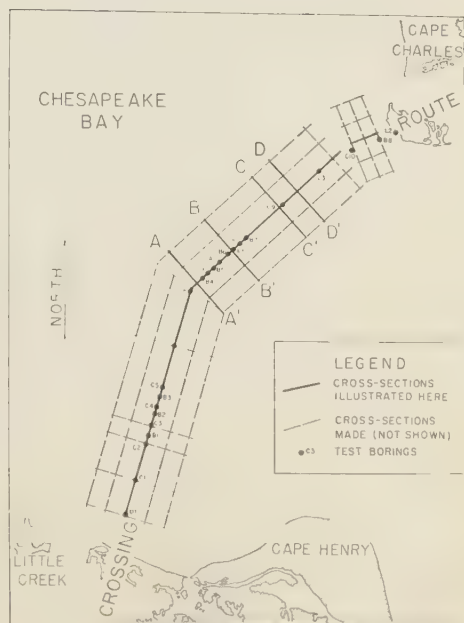


Fig. 1. Map of southern end of Chesapeake Bay showing proposed highway crossing, borehole locations, and survey area.

¹Lamont Geological Observatory Contribution No. 396.

though currents indicated in the current tables for the area are said to be moderate, they are locally strong and, under adverse wind conditions, become appreciable.

The shores of the area are low and have few prominent natural features suitable for landmarks. Even under the best of conditions there are many parts of the survey area from which the shorelines are not visible.

The area surveyed is entirely within the Coastal Plain province, and the crystalline basement rocks lie at depths of the order of several thousand feet. The upper 400 ft, which is of interest to the engineers, is principally Miocene in age [Richards, 1945] with the more recent material of moderate thickness near the top.

The tools. Both the sonic and the navigational equipment are portable and may be installed on small craft without any special alterations. For this survey a 70-ft motorship, *Robin*, was used, and the equipment was installed in a deck-level cabin approximately 6 by 8 ft.

The sonic equipment was a Sub-bottom Depth Recorder (SDR) designed at Lamont Geological Observatory, Columbia University, after an earlier model constructed at the Woods Hole Oceanographic Institution, and constructed by Alpine Geophysical Associates. It consisted of two 4-ft relay racks weighing a total of 250 lb, together with a towed fish, containing the spark source transducer and the hydrophone.

The Raydist navigation system used in this operation consisted of two shore-based stations and duplicate Raydist position indicators mounted on the *Robin*. One set was the newest transistorized Raydist equipment, and the other was an older vacuum-tube system. Each equipment contained a basic electronic unit, two indicators, a strip-chart recorder, and a battery power supply.

The transistorized Raydist equipment requires approximately 2 cu ft of space and weighs 40 lb including battery pack. It can be operated either on batteries or with an external 24-volt supply.

The two shore stations were about 21 miles apart, spanning the mouth of Chesapeake Bay, one being located at Grand View, Virginia (designated the Red station), and the other at Oyster, Virginia, on the Cape Charles penin-

sula (designated the Green station). These were of the 100-watt type, although the 10-watt equipment would have been entirely satisfactory for the purpose.

The methods. Before the SDR survey, Raydist was used to locate the drilling platform for some of the boreholes. Sverdrup and Parry furnished Hastings-Raydist with a detailed boring plan showing the exact spots, in terms of coordinates on the Virginia grid system, at which each boring should be made. The distances to the two shore stations were calculated to the nearest foot for each of them. An individual large-scale plan of each position was then prepared showing the desired location and the Raydist ranges in the immediate vicinity.

The locations were marked by buoys. This was accomplished by navigating the boat along one range line until the desired intersection with the other was reached. The pilot of the vessel was given continuous instructions by the Raydist operator so that the position along the range line was maintained. At the desired location an anchor and marker buoy were dropped. The drilling platform was then moved into position over the anchor.

The experience gained in positioning the boreholes was of great value to the SDR survey. Since the Raydist operator and the pilot had established rapport, it was decided to continue the same operating procedure.

A survey grid was established consisting of tracks along the bridge line and at distances of $\frac{1}{2}$ and 1 mile on either side, together with numerous transverse crossings. The proposed survey line was given to the Raydist operator who then gave the base course to the pilot. The pilot, on the basis of his best judgment as to the nature of the currents and the state of the tide, selected a heading. Through an intercommunication system he was continually given corrections determined by Raydist. Since the survey speed was approximately 5 knots and currents at times were in excess of several knots, it required considerable skill on the part of the pilot to maintain the desired track.

In the shallower parts of the area it was sometimes necessary to make several passes over the same track using different gain settings in order to delineate the upper layers proper

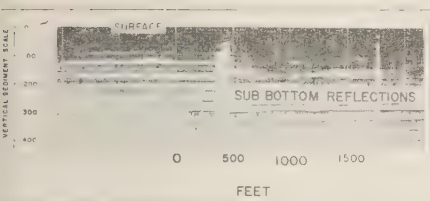


Fig. 2. Six-minute sample of SDR record.

and still to obtain deep penetration. Through use of the Raydist system it was possible to approximate the original lines closely as well as to identify the positions of the boreholes with respect to the surveyed lines. The ability to apply corrections to the ship's heading immediately is a necessity in work of this type. *The results.* The entire survey of the proposed crossing route was completed in about 10 days. Of this time the Raydist operations involved in positioning the 24 open-water boreholes required 17 days whereas the SDR survey required only 7 days including installation, calibration, operation over some 200 miles of track, and removal of equipment.

The over-all positioning accuracy of the boreholes, including maneuvering the *Robin*, dropping the buoy, moving the platform into position, and making it secure, was about ± 25 ft from sites up to 8 miles from land.

Some of the results of the SDR survey are shown in Figures 2 and 3. Figure 2 illustrates a minute segment of some 40 hours of records. Along this section the gain was reduced every

2 or 3 minutes to delineate the bottom and to determine possible shallow reflecting horizons. In other parts of the survey several passes were made over the same tract to accomplish the same result. This problem has now been reduced by means of a time-variable gain arrangement.

Figure 3 illustrates a composite section along the bridge line. Where several passes were made the results from all have been included in the section. In parts of the area it was difficult to gain deep penetration, owing in part to a hard bottom and in part to the presence of organic or gas-filled material. In the former case most of the energy went into multiple reflections, and in the latter absorption of acoustic energy was very high.

Since the parameter measured by the SDR is time, it is necessary to know the velocity structure of the sediments in order to convert to depth. This was done by correlating the velocity-porosity data of *Nafe and Drake* [1957] and *Sutton, Berckhemer, and Nafe* [1957] with the borehole data. The principal reflecting horizons correlated very well with depths in the boreholes at which large changes in porosity occurred or at which the drilling rate showed marked changes.

It turned out to be possible to make an independent check of the accuracy of the navigation from the SDR data. In the course of the drilling operations the bottom sediments were sufficiently disturbed that anomalous reflections were observed on the SDR records at the borehole positions.

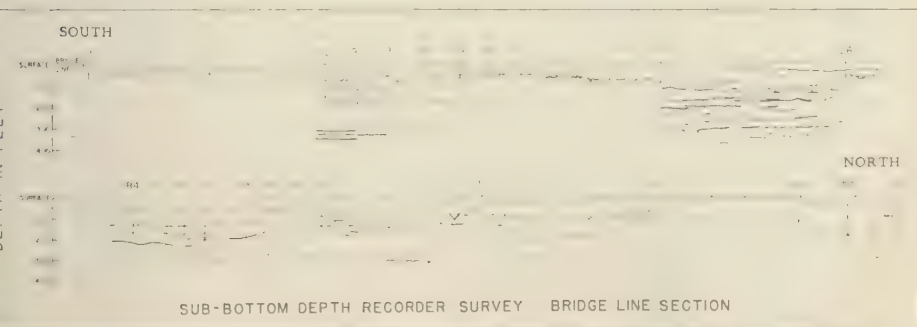


Fig. 3. Composite plot of reflection data along the line of the proposed highway crossing.

Conclusions. The value of high-precision, immediately available navigation for a survey of this type is obvious. Through the use of Raydist the operations could be carried out where desired, and the subsequent reduction of the great volume of SDR data was greatly simplified. The portability of the transistorized apparatus allows it to operate on vessels with limited space and electrical power. Almost any vessel that can safely navigate the area to be surveyed can be used with this equipment.

REFERENCES

- Nafe, J. E., and C. L. Drake, Variations with depth in shallow and deep water marine sediments of porosity, density and the velocities of compressional and shear waves, *Geophysics*, **22**, 1019-1024, July 1957.
- Richards, H. G., Subsurface stratigraphy of the Atlantic Coastal Plain between New Jersey and Georgia, *Bull. Am. Assoc. Petrol. Geologists*, **29**, 885-955, 1945.
- Sutton, G. H., H. Berckhemer, and J. E. Nafe, Physical analysis of deep sea sediments, *Geophysics*, **22**, 779-812, 1957.

Further Evidence of a Solar Corpuscular Influence on Large-Scale Circulation at 300 Mb

NORMAN J. MACDONALD AND WALTER ORR ROBERTS

*High Altitude Observatory
University of Colorado
Boulder, Colorado*

Abstract. Statistical evidence from three successive winter half-years strongly indicates that, when the earth is bombarded by unusually intense solar corpuscular emission, certain troughs in the 300-mb circulation are subsequently amplified. The troughs so affected enter or are formed in the Gulf of Alaska-Aleutian Islands area on the second, third, or fourth day after the start of the corpuscular increases. The trough amplifications maximize a variable number of days later, which may explain why the result was not apparent to earlier workers. The result has been found independently, at about the same level of significance, in the data each of the three winters treated separately. For the data of three half-years grouped together the probability of such a strong chance association is less than 10^{-4} .

Introduction. Earlier results of our work suggested the surprising conclusion that large increases of solar corpuscular emission, signaled by geomagnetic disturbances, were followed in a statistically significant number of instances by changes in the large-scale circulation at the 300-mb level [Woodbridge and others, 1959; Macdonald and Woodbridge, 1959]. These previous results show that for the winter half-years 1956-1957 and 1957-1958 the troughs in the 300-mb wave patterns which moved into or first formed in the Alaska-Aleutians area 2 to 4 days after a geomagnetic disturbance were more likely to have, at full development some 5 to 7 days later, a larger amplitude (greater cyclonic curvature) than troughs appearing in the same area at other times.

In this paper we have adopted what we believe to be improved criteria for defining the arrival of a cloud of solar corpuscles, and have used these criteria in analyses embracing the previously examined periods and also the more recent 1958-1959 winter period.

Solar corpuscular indices. The hypothesis underlying our work is that solar corpuscular emission, when it impinges on the earth in or near the belt of maximum auroral activity, affects meteorological phenomena in a significant number of instances. We have sought, therefore, to use an index that sensitively but objectively selects the beginnings of periods of heightened

solar corpuscular emission in high latitudes and thus defines 'key days' for the study of subsequent meteorological phenomena.

In the earlier analyses we utilized indices based on specified rates of sharp rise in the daily Cheltenham (Maryland) index of geomagnetic disturbance, A_{CH} . With such indices, however, we ran into difficulties when we tried to apply the same numerical criterion of disturbance over the three successive winter periods; the range and means of the A_{CH} index varied widely from year to year, giving us highly different numbers of key days in each of the three periods. To obtain reasonable statistical treatments it was therefore necessary to use different A_{CH} definitions of key dates in each period—a highly unsatisfactory procedure.

For the present study, therefore, we have used a definition of key dates not subject to this difficulty. In this analysis a day was listed as a 'corpuscular (Cp) key day,' if it fell into either or both of the two following categories: (1) It was the day on which an isolated sudden-commencement magnetic storm (Sc) began at Cheltenham, Maryland, or (2) it was an isolated day, or the first day of successive days, on which a strong auroral display was reported from the University of Saskatchewan.¹ These

¹ As reported in a *Preliminary Report of Solar Activity, Series TR*, published each week by the High Altitude Observatory.

TABLE 1. Trough-Distribution Data, Winter Half-Years

	1958- 1959	1957- 1958	1956- 1957
Number of troughs appearing in Gulf of Alaska-Aleutians area	70	63	54
Mean I_t for all troughs at full development	0.600	0.770	0.675
Median I_t for all troughs at full development	0.510	0.550	0.550
Number of troughs dissipating before reaching 0° longitude (percentage of all troughs in parentheses)	28(40)	34(54)	18(33)

criteria not only give us reasonably similar numbers of key dates in each winter period but also seem likely to be a more reliable indicator, at least in years of high solar activity like these, of the outstanding solar corpuscular streams striking the earth. Sudden-commencement magnetic storms are generally considered to be the specific result of the impact on the geomagnetic field of a sharp wave front of solar particles. Strong auroras, moreover, are thought to be an immediate response of the atmosphere to the collisions with the incident particle clouds, even though there are many gaps in the present theory, particularly with respect to the role of 'trapped radiation' belts.

Application of the new criteria does not greatly alter the selection of key dates previously reported for 1956-1957 and 1957-1958. To the degree that it does, the significance of

the associations are improved or unchanged, shown below.

Statistics of 300-mb trough development
For the winter half-year from October 1956 through March 1959, we analyzed the life histories of all detectable 300-mb troughs in exactly the same manner as in the earlier work [Woodbridge and others, 1959]. Meteorologists (not the authors) with no prior knowledge of geomagnetic or auroral activity for the period under study selected the days of first appearance of all detectable troughs that moved into or formed in the Gulf of Alaska-Aleutian Islands area. Most of these, incidentally, moved into the area rather than being formed there. The area (called 'target area' in this paper) was defined as that north of 40°N, west of 120°W and east of 180°.

These meteorologists calculated the trough index I_t for 00 UT of each successive day of the evolution of each trough. The analysis was done from 300-mb-level weather maps of the western half of the northern hemisphere produced by the Department of Transport of Canada. The trough index, described earlier [Woodbridge and others, 1959], measures the amplitude of trough systems. It corresponds to the ratio of the depth to width of the trough at the 30,400-foot and 29,200-foot height contours of the 300-mb pressure surface. Each trough was followed until it dissipated or until it moved off the map east of 0° longitude.

Some of the characteristics of the distribution of troughs for the winter half-year period (October through March) are listed in Table 1. As shown, each of the successive three half-years displayed a greater number of troughs than its predecessor.

The numbers of troughs found within the

TABLE 2. Numbers of Troughs in Indicated Size Classes

	1958-1959		1957-1958		1956-1957	
	I_t Class Limits	No.	I_t Class Limits	No.	I_t Class Limits	No.
Large troughs	≥0.630	24	≥0.850	22	≥0.750	19
	<0.630		<0.850		<0.750	
Medium troughs	>0.400	23	>0.450	19	>0.450	16
	≤0.400	23	≤0.450	22	≤0.450	19

TABLE 3. Number of Troughs That Dissipated

Trough Size	Total Number of Troughs	Number Dissipated	Percentage Dissipated
Maximum development			
Large	65	22	31
Medium	58	18	31
Small	64	40	62

indicated class intervals of trough index I_t for the three periods are shown in Table 2. The class limits were chosen so as to produce approximately equal numbers of troughs in each of three size classifications for each half-year period.

The statistics of the 80 troughs that dissipated, as a function of the size class, for all three inter periods combined are shown in Table 3. It is apparent that a larger percentage of small troughs dissipated than large or medium sized troughs.

Relation of trough development to solar corpuscular emission increases. When the C_p index of solar corpuscular emission was used to define the onset of a solar corpuscular increase, and thus to select the key dates after which to analyze trough development, the results of Tables 4 and 5 were obtained.

In these tables we show in column 2 the numbers of troughs of different ultimate sizes that moved into or formed in the Gulf of Alaska-Aleutian Islands area on the second, third, and fourth day after the arrival of a solar corpuscular cloud, as indicated by the C_p

TABLE 4. Numbers of Troughs of Various Sizes after C_p Dates and Other Dates, 1956-1957, October through March

(1) Trough Size	(2) After C_p Key Dates	(3) Not After C_p Key Dates	(4) Total
Large	11 (6)*	8 (13)	19
Medium	3 (6)	13 (10)	16
Small	4 (6)	15 (13)	19
Total	18	36	54

* Numbers in parentheses indicate random expectation.

TABLE 5. Numbers of Troughs of Various Sizes after C_p Dates and Other Dates, 1957-1958, October through March

(1) Trough Size	(2) After C_p Key Dates	(3) Not After C_p Key Dates	(4) Total
Large	12 (6)*	10 (16)	22
Medium	3 (5)	16 (14)	19
Small	2 (6)	20 (16)	22
Total	17	46	63

* Numbers in parentheses indicate random expectation.

criteria defined above. Column 3 shows the numbers of troughs in these same size classes first appearing in the target area on any other days. The random expectations are shown in parentheses in the tables. Both tables clearly exhibit an association of C_p key dates with larger trough sizes. These tables are significant at the 1 per cent level; i.e., for each table the probability of a chance association as strong as is shown is less than or equal to 0.01, as computed from a chi-square test. These two tables are essentially the same as those previously published, based on the A_{cu} criteria of solar corpuscular emission increases [Macdonald and Woodbridge, 1959]. They exhibit the same or a slightly more significant association.

To test the association found for the earlier two winter periods we carried out a new analysis on entirely independent data. We treated a new half-year, 1958-1959, by exactly the same methods used to obtain the results shown

TABLE 6. Numbers of Troughs of Various Sizes after C_p Dates and Other Dates, 1958-1959, October through March

(1) Trough Size	(2) After C_p Key Dates	(3) Not After C_p Key Dates	(4) Total
Large	11 (6)*	13 (18)	24
Medium	2 (5.5)	21 (17.5)	23
Small	4 (5.5)	19 (17.5)	23
Total	17	53	70

* Numbers in parentheses indicate random expectation.

TABLE 7. Numbers of Troughs of Various Sizes after *Cp* Dates and Other Dates, Three Winter Half-Years

(1) Trough Size	(2) After <i>Cp</i> Key Dates	(3) Not After <i>Cp</i> Key Dates	(4) Total
Large	34 (18)*	31 (47)	65
Medium	8 (16)	50 (42)	58
Small	10 (18)	54 (46)	64
Total	52	135	187

* Numbers in parentheses indicate random expectation.

in Tables 4 and 5. The results are shown in Table 6.

The association of corpuscular increases with larger troughs appears to be of exactly the same nature as that found in the earlier winter periods and at approximately the same level of significance (about 1 per cent).

Combining the results of all 18 months, we obtain Table 7. This table shows a clearly defined association of large troughs with increases of solar corpuscular emission. The random expectations are shown in parentheses. The chance

of finding, by random association, a relationship as strong as this is well under 10^{-6} .

This appears to us to be reasonably conclusive evidence that during these years a real solar-weather effect was at play. The evidence suggests, moreover, that the effect was not simply one that operated in a small though significant percentage of cases. The effect was rather one that appears to have operated for a substantial fraction of all large troughs, and therefore it merits intensive further research.

Average size and dissipation tendencies of troughs following corpuscular key dates. Figure 1 displays, for the three winter half-years, the average trough amplitudes on the days following the day of first trough appearance in the target area. Troughs were separated into two classes according to whether they were or were not associated with *Cp* key dates. It is clear that troughs associated with *Cp* key dates developed into appreciably higher average trough amplitudes on the days following their first appearance in the target area. However, examination of individual trough characteristics shows that the number of days from first trough appearance to maximum trough amplitude fluctuated over substantial limits from trough to trough. Study of individual troughs also revealed that they tended to reach maximum development over different geographical regions. As a consequence, efforts to find this particular solar-weather relation might have failed if attention had been confined to study of trough maxima in fixed geographic locations or at fixed time-lags after *Cp* key dates.

On the other hand, it is not unreasonable to expect that, if there is a real solar effect, both the geographical location of the maximum cyclonic curvature for any given trough system and the time of this occurrence might be dependent on the state of the circulation at the time of the solar corpuscular increase and not solely on the solar impulse itself.

When we had discovered the above results we felt that there might be a systematic difference in the dissipative tendencies of the troughs that remained small, even though they followed *Cp* key dates, as compared with others that remained small at their maxima. We expected that the troughs preceded by *Cp* dates might show less tendency to dissipate. Table 7 shows that this is indeed the case. The associ-

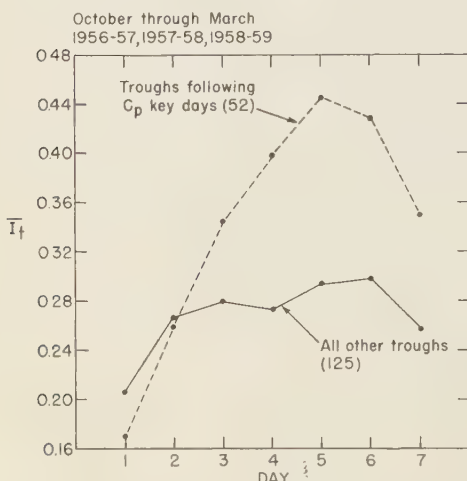


Fig. 1. Average trough amplitude I_t for the seven days after first appearance in the target area for all troughs that did not move off the maps (past 0° longitude) during the seven days. Dotted curves for troughs that first appeared in target area on second, third, and fourth days after *Cp* key dates. Solid curves, all others.

TABLE 8. Dissipative Tendencies of Small Troughs in Three Winter Half-Years: 1956-1957, 1957-1958, 1958-1959

Troughs	After C_p Key Dates	Not After C_p Key Dates	Total
Dissipating	7 (4)*	17 (20)	24
Dissipating	3 (6)	37 (34)	40
Total	10	54	64

* Numbers in parentheses indicate random expectation.

is significant below the 5 per cent level by chi-square test.

Choice of time lag between C_p date and first trough appearance. The final step in the present analysis was to examine objectively the question whether we had made an optimum choice in looking only at troughs first appearing in the target area on the second, third, and fourth day after the C_p key dates of solar corpuscular increase. The conclusions derived from the seasons 1957-1958 and 1958-1959 are summarized in Figure 2. This is an independent test of the thesis developed from a qualitative inspection of the 1956-1957 weather maps to the effect that days 2, 3, and 4 after the corpuscular increase were the significant dates for first appearance of the troughs in the key target area. It confirms objectively the hypothesis that there is no other comparably significant time-lag for the association, at least between the fourth day before to the tenth day after the C_p dates. If several periods had exhibited equal significance, the reality of the whole solar-weather relationship would have fallen into question, of course. From this result we can conclude that it is *only* those troughs that appear in the target area from 2 to 4 days after corpuscular rises that exhibit the apparent solar-weather effect.

Discussion of results. Although the present study presents strong evidence that there is a real solar corpuscular influence on the development of waves in the westerly winds at 300-mb elevations, no physical mechanisms have been offered to explain what must simply be regarded as an empirical-statistical finding. The results suggest, of course, that some solar influence associated with auroral and geomagnetic



Fig. 2. The probability P that the observed association of large troughs with C_p key dates could occur by chance. P is shown for successive three-day periods centered on days before (—) or after (+) C_p key days. Results combine data for the 1957-58 and 1958-1959 periods. Confidence limits from chi-square test are shown by the dashed horizontal lines.

storm phenomena affects the stability of large-scale wave systems. But the amount of energy generally assumed to be present in the entire cloud or clouds of arriving solar corpuscles responsible for a magnetic storm and aurora is many orders of magnitude smaller than the differences of kinetic energy between small and large trough systems. Where, then, can the energy for the developments be found? Moreover, how can small amounts of solar energy absorbed high in the atmosphere, probably above 50 km, affect large-scale wind motions below 15 km?

At this time, only qualitative speculations are possible, and we sincerely hope that other workers will attack the problem afresh and with vigor. Perhaps, as London [1956], has suggested heating at the top of the atmosphere can produce ascending motions in the stratosphere that, in turn, make the lower atmos-

phere more unstable or aid in the vertical exchange of momentum.

But a 'brute force' heating mechanism may not be the most likely physical process. We have postulated elsewhere [Macdonald and Roberts, 1958] that an explanation of the mechanisms for this kind of solar effect on planetary wind systems may involve atmospheric nuclei produced by the solar corpuscular impulse. These may, we speculate, modify the effectiveness of condensation and sublimation processes at lower atmospheric levels. Such processes might, in turn, affect the large-scale circulation, as the work of Aubert [1957] has suggested. Perhaps atmospheric nuclei are produced or rendered effective as freezing nuclei by an increase in the bremsstrahlung X-ray radiation that penetrates to stratospheric levels as a secondary effect of aurora-producing solar corpuscles. Such X-rays have been observed from balloons by Winckler and others [1958], who have found them as low as 60,000 feet above sea level during strong auroras.

Strengthening the suggestion that atmospheric nuclei may be involved is the work of Barber [1955]. He found that on days after magnetic disturbances there were indications of greater light-scattering in the zenith sky. The scattering, he surmised, resulted from a greater abundance, after magnetic disturbances, of atmospheric particles large in size compared with the wavelength of light.

To pursue this possibility further, the High Altitude Observatory plans to conduct studies of the brightness of atmospheric solar halos using newly developed techniques [Newkirk, 1956; Deirmendjean, 1959]. Such experiments would be of particular value if they could be done in the Gulf of Alaska-Aleutian Islands area, which appears from this study to be particularly significant. The possibility of small-scale, high-altitude nuclear experiments on meteorologically chosen dates is also an attractive one. Such nuclear blasts might release radiations capable of producing an atmospheric effect analogous to that of a natural cloud of solar corpuscles and might conceivably have large-scale meteorological consequences.

The results reported here are probably not independent of the suggested effects reported by Arai [1958], by Chernov and Khistyakov [1956], and by others. We have discussed some

of these possible inter-connections elsewhere [Macdonald and Roberts, 1959].

Acknowledgments. We wish to thank R. F. Brun and Karen M. Lowman for their work in computing the trough indices and for determining the starting dates of trough systems. We thank Herbert H. Riehl for constructive criticism and helpful suggestions at various stages of work. Finally, we express gratitude to P. D. Taggart-Cowan and Andrew Thomson of the Canadian Department of Transport for the 300 map analyses they have sent us for the past two years.

This research was supported by the privately sponsored Institute for Solar-Terrestrial Research of the High Altitude Observatory.

REFERENCES

- Arai, Y., Characteristics of long waves in weather patterns related to solar activity, *J. Meteorol. Japan*, **36**, 46-54, 1958.
- Aubert, E. J., On the release of latent heat as a factor in large scale atmospheric motions, *Meteorol.*, **14**, 527-542, 1957.
- Barber, D. R., Changes in brightness, polarization and colour of the zenith day sky accompanying geomagnetic activity, *J. Atmospheric Terrest. Phys.*, **7**, 170-172, 1955.
- Chernov, B., and V. Khistyakov, Halos and general circulation of the atmosphere, *Bull. Union Astron. Geophys. Soc.*, no. 19, Moscow, USSR, Sept. 1956.
- Deirmendjean, D., Theory of the solar aureole, *Ann. géophys.*, **16**, 218-240, 1959.
- London, J., Solar eruptions and the weather, *Trans. N. Y. Acad. Sci.*, **19**, 138-146, 1956.
- Macdonald, N. J., and D. D. Woodbridge, Relation of geomagnetic disturbances to circulation changes at the 30,000-foot level, *Science*, **165**, 638-639, 1959.
- Macdonald, N. J., and W. O. Roberts, The relationship of 300 mb circulation changes to geomagnetic disturbances from 1952 to 1958, *Inst. Solar Terrest. Research, High Altitude Observatory, Univ. Colo.*, Rept. 6, 1958.
- Macdonald, N. J., and W. O. Roberts, Further evidence of a solar corpuscular effect on weather disturbances at high altitudes, *Inst. Solar Terrest. Research, High Altitude Observatory, Univ. Colo.*, Rept. 7, 1959.
- Newkirk, G. A., Photometry of the solar aureole, *J. Opt. Soc. Am.*, **46**, 1028-1037, 1956.
- Winckler, J. R., L. Peterson, R. Arnoldy, and J. Hoffman, X-Rays from visible aurorae at Monticello, *Phys. Rev.*, **110**, 1221-1231, 1958.
- Woodbridge, D. D., N. J. Macdonald, and T. Pöhrte, An apparent relationship between geomagnetic disturbances and changes in atmospheric circulation at 300 millibars, *J. Geophys. Research*, **64**, 331-341, 1959.

(Manuscript received October 23, 1959.)

The Ring Current and the Outer Atmosphere

SYUN-ICHI AKASOFU

*Geophysical Institute, University of Alaska
College, Alaska*

Abstract. It is shown that hydromagnetic propagation will enable the earth's field at ground level to be affected with a lag of less than 1 minute by changes in electric currents far above the earth—for example, by a changing ring current with a radius a few times that of the earth. It is shown that a system of trapped high-speed charged particles, such as those in the Van Allen belts, involves electric current flow; the deciding factor is the curvature of the guiding lines of the geomagnetic field. Thus an enhancement of the Van Allen radiation belts will decrease the horizontal geomagnetic component. Such a change is observed during the main phase of a magnetic storm. A suggestion is made about the origin of trapped high-energy particles from the analysis of two large magnetic storms during the IGY.

INTRODUCTION

The decrease of horizontal intensity during the main phase of a magnetic storm has been attributed to the growth of a westward electric current around the earth (cf. *Chamberlain*, 1958, p. 168, for a review of the subject up to date). Until recent years the ring current was regarded as flowing *in vacuo*, and the changes in its field were supposed to reach the earth at the speed of light, unaffected by the ionosphere except for conductive shielding [*Ashour Price*, 1948; *Sugiura*, 1949; *Ferraro*, 1954].

Nowadays the atmosphere is believed to extend over the region (out to a few earth radii) in which, since the work of *Chapman and Ferraro* [1951], the ring current has been supposed to flow. The whistler phenomenon as interpreted by *Storey* [1954] indicated that ionized gas must also be present in that region. This new theory has generated a controversy about the propagation of magnetic changes through the ionized atmosphere to the earth's surface. *Hines* [1958] and *Hines and Storey* [1958] concluded that the propagation could be achieved in 10 minutes or so by hydromagnetic waves.

The first part of this paper develops this conclusion in fuller detail and with slightly different numerical conclusions. In the second part the electric current distribution associated with the Van Allen radiation belts is considered. The third part is an analysis of two large magnetic storms during the IGY.

PROPAGATION OF FIELD CHANGES IN AND THROUGH THE IONOSPHERE

Ashour and Price [1948] and *Sugiura* [1949] discussed the electric currents induced in the ionosphere by changing magnetic fields at higher levels. The ionosphere was for this purpose treated as a rigid, concentric conducting shell. Such a treatment is summarized by *Ferraro* [1954]. He shows that a periodic disturbing field outside the shell, with potential given by

$$\psi = E_0 \cos \alpha t \, r \cos \theta \quad (1)$$

gives rise to currents such that the resultant field within the shell has the potential

$$\psi' = \operatorname{Re} \frac{E_0 r \cos \theta}{1 + i\alpha\lambda} e^{i\alpha t} \quad (2)$$

(where Re means 'the real part of'). Here $\lambda = 4\pi aK/3$, and a denotes the radius of the shell, K its integrated conductivity (emu).

For an aperiodic external field given by

$$\psi = f(t) E_0 r \cos \theta \quad (3)$$

when

$$f(t) = 0 \quad \text{for } t < 0$$

$$f(t) = 1 \quad \text{for } t \geq 0$$

we have similarly

$$\psi' = E_0 \frac{(1 - e^{-\tau}) - \alpha\lambda(1 - e^{-\tau/\alpha\lambda})}{1 - \alpha\lambda} r \cos \theta \quad (4)$$

where

$$\tau = \alpha t \quad \text{and} \quad \alpha\lambda \neq 1$$

Equations 2 and 4 indicate that the ionosphere modifies the field within it. If K is infinite the interior is completely shielded from external changes caused by currents (of density J).

However, such a treatment of the problem is inadequate because the ionospheric shell is not rigid; it is a partially ionized gas, a hydromagnetic fluid. When this is taken into account the conclusions are much changed.

The fundamental hydromagnetic equations in the presence of a magnetic field \mathbf{H} are:

$$\frac{\partial \mathbf{H}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{H}) + \frac{c^2}{4\pi\sigma} \nabla^2 \mathbf{H} \quad (5)$$

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{F} + \mathbf{J} \times \mathbf{H} - \nabla p \quad (6)$$

where \mathbf{v} denotes the velocity of the medium, p the pressure, \mathbf{F} the external force per unit volume, and σ the electrical conductivity. The treatment associated with equations 1 to 4 assumes that $\mathbf{v} = 0$. In this case (5) becomes

$$\partial \mathbf{H} / \partial t = (c^2 / 4\pi\sigma) \nabla^2 \mathbf{H} \quad (7)$$

a diffusion equation. A simple dimensional argument shows that its time scale T is of the order

$$4\pi\sigma L^2 / c^2 \quad (8)$$

where L is a characteristic length for the problem. When the medium can move, the ratio (first term/second term) on the right of (5), called by *Elsasser* [1956] the magnetic Reynolds number R_m , is given by

$$4\pi\sigma L v / c^2 \quad (9)$$

For hydromagnetic changes in the far outer atmosphere we may take $L = 5$ earth radii, $\sigma = 2 \times 10^7 \times T^{3/2} = 6 \times 10^{11}$ esu (for $T = 1000^\circ\text{K}$), and \mathbf{v} may there exceed 1 meter/sec*. These $R_m = 2.7 \times 10^3$, so that the second term on the right of (5) can be neglected.

In that case, as *Alfvén* [1950] first showed, equations 5 and 6 lead to the following wave equation for small perturbations h in the general dipole field,

$$\frac{\partial^2 h}{\partial t^2} = \left\{ \frac{H}{(4\pi\rho)^{1/2}} \right\}^2 \frac{\partial^2 h}{\partial x^2}$$

This indicates that the magnetic changes do not diffuse, but are propagated with speed $H/(4\pi\rho)^{1/2}$. *Hines* [1959] and *Hines and Storey* [1958] stressed this point in discussing the argument of *Parker* [1956], which was based on (7) and (8). *Hines* [1953] first derived the dispersion equation (showing the dependence of the phase velocity on the angular frequency) for hydromagnetic waves in a ternary mixture (positive ions, electrons, and neutral atoms). Later it was discussed by *Dungey* [1954a, b] and *Akasofu* [1956] in the case of the ionosphere. Recent estimates of the speed of such waves have been made by *Francis and others* [1958] using new data for the electron density distribution in the outer atmosphere; they indicate that an approximate value over most of the period is 1000 km/sec. At this speed, changes of the ring-current field would reach the earth's surface in about 30 seconds, instead of in 80 days, (8) would imply. It is also much shorter than former estimates: *Akasofu* [1956], 5 minutes; *Hines and Storey* [1958], 10 minutes. This important difference results from the fact that ions in the upper atmosphere can move in response to changes in the external field, if the characteristic time of these changes is less than the gyromagnetic period $2\pi mc/eH$; in the outer layer and at 5 earth radii these periods are 3.5×10^{-2} sec and 2.6×10^{-1} sec. Owing to this ionic motion the magnetic field can change quickly, even in the region where the magnetic pressure $H^2/8\pi$ much exceeds the gas pressure. The theory of propagation of hydromagnetic waves in the ionosphere is an extension of the magneto-ionic theory to the case where the ions can easily move in response to field changes [*Åström*, 1950; *Hines*, 1953; *Dungey*, 1958].

RING CURRENT

In considering the ring current, it is not possible to start from the observed fact of the Van Allen belts surrounding the earth, as revealed by United States and Soviet satellites [*Van Allen and Frank*, 1959]. *Singer* [1957] with notable insight, at least two years before their discovery, inferred the probable existence and nature of charged particles around the earth

* See page 540, Eq. (32).

Following *Alfvén's* [1950] study of motions of charged particles around a dipole field. He pictured the spiraling of high-speed positive and negative charges around the lines of geomagnetic force, back and forth across the equator, between relatively low levels at high latitudes, where the motion along the lines of force is slowed and reversed. He saw also that the positive particles would drift westward around the earth and the negative charges eastward, owing to the gradient of the earth's magnetic field there. He identified this process with the ring current, though it is shown in the present paper that this drift motion does not contribute to the ring current.

Fundamental equations. The motion of the particles is decidedly complex, even disregarding the mutual action between them. *Alfvén* [1950] showed how to calculate a 'smoothed out' position of the path of such particles, using a perturbation method; it is applicable when the time change and space change in the magnetic field traversed by a particle during one gyration is small compared with the average field intensity along the particular spiral. *Alfvén* introduced the concept of the guiding center of a particle; its path gives a general idea of the course of the particle, omitting the details of the gyrations. For protons with energy of 100 Mev at the distance of 5 earth radii, the radius of its spiral motion is of the order of 180 km—small in comparison with the size of the belts and with the characteristic distance for a field change of a few per cent.

When there is an assembly of particles, positive and negative, in the same region, it cannot be assumed that their motions will be similar to those of single particles traveling alone; their motions cannot be independent. But the present paper is concerned with a steady state, in which limited space in the outer atmosphere (the Van Allen belts) is filled by equal numbers of protons and electrons.

Parker [1957], considering a plasma in a magnetic field, distinguished three separate components of electric current associated with the particle motions:

1. Circulating current:

$$\mathbf{i}_C = (c/8\pi p_m) \mathbf{H} \times \{ \nabla p_n - \frac{1}{2}(p_n/p_m) \nabla p_m - (\mathbf{H} \cdot \nabla) \mathbf{H} / 8\pi \} \quad (11)$$

2. Drift current:

$$\mathbf{i}_D = (c/8\pi p_m) \mathbf{H} \times \{ \frac{1}{2}(p_n/p_m) \nabla p_m + (p_s/p_m)(\mathbf{H} \cdot \nabla) \mathbf{H} / 8\pi \} \quad (12)$$

3. Polarization current:

$$\mathbf{i}_P = (\rho c/H^2) \mathbf{H} \times (d\mathbf{v}/dt) \quad (13)$$

where

p_s = the total pressure component parallel to \mathbf{H} , namely mnw_s^2 .

p_n = the total pressure component perpendicular to \mathbf{H} , namely $\frac{1}{2}mnw_n^2$.

w_s = the velocity component parallel to \mathbf{H} .

w_n = the velocity component perpendicular to \mathbf{H} .

m = the mass of each charged particle.

n = the number density of charged particles.

H = the intensity of the earth's magnetic field.

$p_m = H^2/8\pi$.

$\mathbf{v} = c(\mathbf{E} \times \mathbf{H})/H^2$.

We assume that the number densities of protons and electrons are the same (electrical neutrality). The current that appears in the ionized gas is the sum of these three currents:

$$\mathbf{i} = \mathbf{i}_L + \mathbf{i}_D + \mathbf{i}_P = (c/8\pi p_m) \mathbf{H} \cdot \{ \nabla p_n + [(p_s - p_n)/p_m] \cdot (\mathbf{H} \cdot \nabla) \mathbf{H} / 8\pi + \rho(d\mathbf{v}/dt) \} \quad (14)$$

It should be noticed that the first term in the main bracket of (12), the electric current caused by the gradient of the Earth's magnetic field, is cancelled by the second term in the main bracket of (11). Thus these two terms do not appear in (14).

The drift motion and the ring current. The current \mathbf{i} given by (14) has three parts, corresponding to the three terms in the main bracket. The first of these, ∇p_n , was considered by *Tonks* [1955] in an important pioneer paper on current in plasmas. Later *Schmidt* [1958] elaborated on Tonk's discussion. They found that in general the plasma tends to be repelled from the more intense region of the field, and that a pressure gradient is set up which produces an electric current flow. However, this term seems likely to be of minor importance, though the distribution of high-energy particles in the Van Allen belts is not yet known.

TABLE 1. The Drift Speed (cm/sec) of Protons whose Equatorial Speed is 2×10^8 cm/sec at Various Latitudes along the Line of Force which Crosses the Equatorial Plane at 5 Earth Radii

Latitude						
0°	10°	20°	30°	40°	50°	60°
$+7.68 \times 10^6$ *	$+6.14 \times 10^6$	$+3.94 \times 10^6$	$+2.03 \times 10^6$	$+5.70 \times 10^4$	$+1.75 \times 10^4$	$+6.72 \times 10^3$
7.62	6.03	3.41	1.50	5.37	1.45	6.17
7.38	6.00	3.39	1.48	5.20	1.31	-1.25
7.55	5.96	3.36	1.46	5.00	1.08	
7.45	5.88	3.28	1.47	4.47	5.58×10^3	
7.33	5.82	3.24	1.36	4.13	2.40	
7.01	5.47	2.95	1.11	2.00		
6.34	4.86	2.44	6.95×10^4	-1.69		
4.80	3.47	1.27	-2.65×10^6			
2.92	1.76	-1.17				

* Plus sign means westward and minus sign eastward drift.

Let θ_s denote the angle between H and the velocity vector w_s of a particle when it crosses the equator at radius r_s (where $H = H_s$). Thus

$$w_n = w_s \tan \theta_s = w_s \sin \theta_s \quad (24)$$

$$w_n = w \sin \theta, \quad w_s = w \cos \theta \quad (25)$$

When a particle moving along a line of force crosses a point P where the intensity of the magnetic field is H , the angle θ is given [Spitzer, 1966, p. 11] by

$$\sin^2 \theta = (H/H_s) \sin^2 \theta_s \quad (26)$$

Consequently, $w_s = 0$ and $\theta = 90^\circ$ at latitude (the 'mirror' latitude)

$$= H_s \frac{(1 + 3 \sin^2 \varphi_0)^{1/2}}{\cos^6 \varphi_0} = \frac{H_s}{\sin^2 \theta_s} \quad (27)$$

The factor $w_s^2 - \frac{1}{2}w_n^2$ becomes zero when θ becomes $\tan^{-1} (2)^{1/2}$ at the latitude φ' given by

$$\frac{1 + 3 \cos^2 \varphi'}{\cos^6 \varphi'} = \frac{\sin^2 (\tan^{-1} (2)^{1/2})}{\sin^2 \theta_s} = \frac{0.665}{\sin^2 \theta_s} \quad (28)$$

At still higher latitudes, up to φ_0 , the protons drift eastward.

Consider a particle that crosses the equatorial plane at a point P_s at distance r_s from the center of the earth. Let subscript e denote reference to the value of H or of the speed and other characteristics of the motion of the particle at this

point. Values of θ , w_s , w_n , r , and H at other points along the path of the parabolic can be obtained in terms of H_s , w_s , θ_s from equations 26, 25, 18, and 20. For example, Table 1 gives, for several values of θ_s , the drift velocities of protons at several points along a line of force, for the initial values 3.2×10^8 cm (5 earth radii) for r_s and 2×10^8 cm/sec (~ 20 kev) for w_s . The table suggests that the main current flows in the lower latitudes, though it critically depends on the anisotropy of the velocity distribution of the particles. Dessler and Parker [1959] have discussed this problem in essentially the same way.

The number density of particles required to produce the observed change of the horizontal intensity of the earth's field at the surface of the earth depends on the energy distribution of the particles. This is not yet known. Therefore only a tentative model is presented in the present paper.

Let C denote the total ring current and Q and b the linear radius of the cross section (treated as circular) of the ring, respectively. The change of the horizontal component ΔH of the earth's field is given [Chapman, 1952] by

$$C = NeV = cQ\Delta H/2\pi \quad (29)$$

where N denotes the average number density of charges times the area πb^2 . Taking $Q = 5 \times 6.38 \times 10^8$ cm and $\Delta H = 100\gamma 10^{-3}$ gauss, these give $C = 1.53 \times 10^{18}$ esu ($= 5.1 \times 10^4$ amperes). Thus N is given by

$$N = C/eV = 3.19 \times 10^{25}/V \text{ per cm} \quad (30)$$

The drift velocity of 20-kev protons at 5 earth radii is given by (23); it is 7.34×10^8 cm. This yields $N = 4.35 \times 10^{19}$ /cm. The radius b of the ring may be of the order 6.38×10^8 cm/2, so that the number density n is 136/cm³. The number density of 100-kev electrons will be 27/cm³ under the same conditions. If b equals the earth's radius, these values of n are reduced respectively to 27 and 7.

When the ring current begins to grow, an electric field is induced and the charged particles react to these changes of the electromagnetic fields; that is, the particles move in the electromagnetic field consisting of the earth's dipole field and the magnetic and electric fields generated by themselves. Parker [1958] discussed a ring current flowing in a copper wire, an analogy that led to a discussion between Hines and Parker [1958]. The induced electric field ($\partial \mathbf{H}/\partial t = -\nabla \times \mathbf{E}$) will produce an outward motion of the charged particles, tending to cause expansion of the outer atmosphere. The corresponding reduction of pressure (magnetic plus gas pressure) within the ring will be propagated toward the earth's surface at the hydromagnetic speed $H/(4\pi\rho)^{1/2} \sim 1000$ km/sec, and will decrease the horizontal component of the field. The result, except for a delay of about 30 seconds, will be essentially the same as that derived by Chapman and Ferraro.

The velocity v of expansion of the outer atmosphere is given (cf. equation 5) by

$$\Delta H/\Delta t \sim v \times H_e/L \quad (31)$$

where L denotes a characteristic length. Sugiura and Chapman [1958] show that, for great storms, the maximum change of the horizontal component is reached about 6 hours after the beginning of the main phase and is about 80γ . Therefore, we may take $\Delta H/\Delta t \sim 80\gamma/6$ hours = $3.7 \times 10^{-2}\gamma/\text{sec}$ and $L = 6.38 \times 10^8$ cm. Taking $H_e = 256\gamma$ at 5 earth radii

$$v \sim \frac{(\Delta H/\Delta t)L}{H_e} = 10^4 \text{ cm/sec} \quad (32)$$

ORIGIN OF HIGH-ENERGY PARTICLES

In this paper it is suggested that the Van Allen belt is the seat of the ring current. The

outer belt appears to be fed by solar gas; distribution changed considerably during several months of observation (from December 1958 to March 1959) [Van Allen, 1959].

However, as pointed out by Van Allen [1959], the energy of many of the particles far exceeds the presumed energy of the solar corpuscles inferred from the time lag between outstanding solar events and subsequent magnetic disturbances.

Furthermore, Chapman has pointed out another difficulty in supposing that such high-energy particles originate in a rather uniform solar stream. Chapman and Ferraro [1940] solved particular problems of the motion of a gas assembly of positive and negative particles traveling together into a magnetic field. They showed that such a gas could penetrate more deeply into the field than a single particle of the same mass, speed, and charge traveling alone. The depth of penetration of the gas is determined not merely by the masses, speeds, and charges of the individual particles but also by their number density.

They showed in detail how the energy of a magnetic field, as this is penetrated by a neutralized stream, increases at the expense of kinetic energy. In an illustrative example designed to fit some aspects of the first phase of actual magnetic storms, they showed how an initial speed of 10^8 cm/sec for protons and electrons may be reduced to transverse speeds of 5.5×10^2 cm/sec (protons) and 10^6 cm/sec (electrons) at the deepest penetration. Particles of such low energy may not contribute much to the ring current (See the Note on p. 543).

Several acceleration mechanisms for solar particles in the outer atmosphere have been proposed, but all of them have discussed only rather constant accelerations and constant flow into the auroral zone during the magnetic storms. However, it is shown in another paper [Akasofu, in preparation] that polar magnetic disturbances (negative bays or DS disturbances) associated with active auroral displays are sporadic. They consist of several large impulses which appear abruptly and also subside rapidly. This suggests that the auroral particles (protons and electrons) associated with polar magnetic disturbance and active auroras appear very irregularly. Winckler and others [1958, 1959] have observed sporadic X-ray bursts which

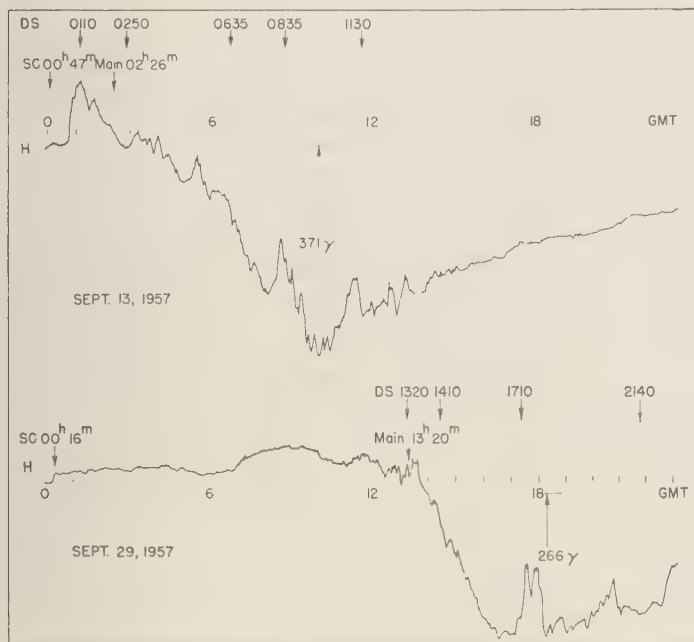


Fig. 3. Horizontal component of magnetograms for two large magnetic storms: *Upper*, September 13, 1957; *lower*, September 29, 1957, from Hermans Observatory (geomagnetic longitude $80^{\circ}3'$, latitude $33^{\circ}35'$), near Capetown. The beginning times of the sudden commencements (SC), the main phases (Main), and the DS's (DS) are also shown.

appeared simultaneously with the DS disturbances.

Both the polar magnetic disturbances and the growth of the ring current may be due to special features, as yet unknown, in the solar stream, allowing abnormally energetic particles to stream to descend into the polar atmosphere and also, in lower latitudes, into the Van Allen belt.

Magnetic storm evidence seems to support this view. For example, *Chapman* [1952] and *Chapman and Chapman* [1958], in their extensive studies of the average properties of magnetic storms, found that the DS disturbance (preeminently polar) and the main storm phase began at about the same time. Two great magnetic storms during September 1957 and interesting individual examples of this approximate coincidence of the onset of both the DS disturbance and the main storm phase, and also of how the interval is from the sudden commencement to this onset, in different storms.

The first storm began at $00^{\text{h}}47^{\text{m}}$ on September 13, the second at $00^{\text{h}}16^{\text{m}}$ on September 29—in both cases, as it chanced, near Greenwich midnight. But in the first storm the onset of the main phase began after $1^{\text{h}}39^{\text{m}}$, whereas in the second the interval was $13^{\text{h}}4^{\text{m}}$. These intervals are derived as the average for 10 equatorial observatories. Figure 3 shows the Hermanus (geomagnetic longitude $80^{\circ}3'$, latitude $33^{\circ}35'$) horizontal force magnetograms for the two storms, and also the initial times of the large DS disturbances, derived from magnetograms for 20 arctic observatories.

Because of the short life of the individual DS disturbances, the special structures in the solar stream would seem to correspond to irregularities in the background stream, of extent, along the stream, of the order of 3 or 4 million km ($1 \text{ hour} \times 1000 \text{ km/sec}$).

The nature of these structures is uncertain. They may be regions in which part of a solar magnetic field is being carried along by the

stream, enabling the stream to hold within itself charged particles of much higher energy than what corresponds to the general forward speed of the stream. Even so, much remains obscure as to the precise way in which such energetic particles cease to belong to the stream and become part of the Van Allen belts.

Thus, it is supposed that the ring current grows by injection of high-energy particles from unknown structures in the solar stream into the Van Allen radiation belt; this enhances the net westward current.

The magnetic field produced by this current system will cause the outer atmosphere to expand. The expansion will be propagated as a finite expansion wave, with a velocity of $H/(4\pi\rho)^{1/2} \sim 1000$ km/sec, toward the earth's surface and will cause a decrease in the horizontal component of the field.

Acknowledgments. The writer wishes to express his sincere thanks to Professor S. Chapman for suggesting this research to him, and for discussion and aid in preparing this paper, and to Dr. C. T. Elvey for his kind encouragement. He is also indebted to Messrs. A. J. Dessler, C. O. Hines, W. Murcray, E. N. Parker, and R. Parthasarathy.

REFERENCES

- Akasofu, S., On the geomagnetic micropulsation, *Rept. Ionosphere Research Japan*, 10, 227-249, 1956.
- Alfvén, H., *Cosmical Electrodynamics*, Oxford University Press, 1950.
- Ashour, A. A., and A. T. Price, The induction of electric currents in a non-uniform ionosphere, *Proc. Roy. Soc. London, A*, 195, 198-224, 1948.
- Åström, E., On the waves in an ionized gas, *Arkiv Fysik*, 2, 443-458, 1950.
- Chamberlain, J. W., *Theories of the Aurora, Advances in Geophysics*, edited by H. E. Landsberg, and J. van Mieghem, Academic Press, New York, 1958.
- Chapman, S., The morphology of geomagnetic storms: An extension of the analysis of DS, the disturbance local-time inequality, *Ann. geofis.*, 5, 481-499, 1952.
- Chapman, S., Theories of the aurora polaris, *Ann. de Geophys.*, 8, 209-255, 1952.
- Chapman, S., and V. C. A. Ferraro, A new theory of magnetic storms, *Terrestrial Magnetism and Atmospheric Elec.*, 36, 77-97 and 171-186, 1931.
- Chapman, S., and V. C. A. Ferraro, The theory of the first phase of a geomagnetic storm, *Terrestrial Magnetism and Atmospheric Elec.*, 45, 245-268, 1940.
- Chapman, S., and V. C. A. Ferraro, The geomagnetic ring-current, I, its radial stability, *Terrestrial Magnetism and Atmospheric Elec.*, 1-6, 1941.
- Dessler, A. J., and E. N. Parker, Hydromagnetic theory of geomagnetic storms, *J. Geophys. Research*, 64, 2239-2252, 1959.
- Dungey, J. W., The propagation of Alfvén waves through the ionosphere, *Penn. State Univ. Rpt.*, 57, 1954a.
- Dungey, J. W., *Electrodynamics of the outer atmosphere*, *Penn. State Univ. Rpt.*, 69, 1954.
- Dungey, J. W., *Cosmic Electrodynamics*, Cambridge University Press, 1958.
- Elsasser, W. M., Hydromagnetic theory, *Rev. Modern Phys.*, 28, 135-163, 1956.
- Ferraro, V. C. A., *Electromagnetic Theory*, University of London Press, 1954.
- Francis, W. E., M. I. Green, and A. J. Dessler, Hydromagnetic propagation of sudden commencements, *J. Geophys. Research*, 64, 1645, 1959.
- Hines, C. O., Generalized magneto-hydrodynamic formulae, *Proc. Cambridge Phil. Soc.*, 49, 2307, 1953.
- Hines, C. O., On the geomagnetic storm effect, *J. Geophys. Research*, 63, 671-682, 1958.
- Hines, C. O., and E. N. Parker, Statement of differences regarding the ring-current effect, *J. Geophys. Research*, 63, 691-692, 1958.
- Hines, C. O., and L. R. O. Storey, Time constant in the geomagnetic storm effect, *J. Geophys. Research*, 63, 671-682, 1958.
- Parker, E. N., On the geomagnetic storm effect, *J. Geophys. Research*, 61, 625-637, 1956.
- Parker, E. N., Newtonian development of the hydromagnetic properties of ionized gases of density, *Phys. Rev.*, 107, 924-933, 1957.
- Parker, E. N., Inadequacy of ring-current theory for the main phase of a magnetic storm, *J. Geophys. Research*, 63, 683-689, 1958.
- Schmidt, G., An investigation on plasma in the external magnetic field, I, *Nuovo cimento*, 19, 55-67, 1958.
- Singer, S. F., A new model of magnetic storms and aurorae, *Trans. Am. Geophys. Union*, 38, 190, 1957.
- Spitzer, L., *Physics of fully ionized gases*, Interscience Pub., Inc., New York, 1956.
- Storey, L. R. O., An investigation of whistling aurorae, *Phil. Trans. Roy. Soc. London*, 247, 113-141, 1953.
- Sugiura, M., Electromagnetic induction in the ionosphere, *Rept. Ionosphere Research Japan*, 3, 65-72, 1949.
- Sugiura, M., and S. Chapman, *Geophys. Int. Univ. Alaska Pub.*, 1958.
- Tonks, L., Particle transport, electric currents and pressure balance in a magnetically immobile plasma, *Phys. Rev.*, 97, 1443-1445, 1955.
- Van Allen, J. A., Lecture in the High Altitude Observatory, September, 1959a.
- Van Allen, J. A., Radiation belts around the earth, *Sci. American*, 200, 39-47, 1959b.

Allen, J. A., and L. A. Frank, Survey of radiation around the earth to a radial distance of 400 kilometers, *Nature*, 183, 430-434, 1959.

Kler, J. R., L. Peterson, R. Arnoldy, and R. Hoffman, X-rays from visible aurorae at Minneapolis, *Phys. Rev.*, 110, 1221-1231, 1958.

Kler, J. R., L. Peterson, R. Hoffman, and R. Arnoldy, Auroral X-rays, cosmic rays, and related phenomena during the storm of February 11, 1958, *J. Geophys. Research*, 64, 597-610, 1959.

Note added in page proof: However, Ferraro [1952] showed that in the front of the stream the electrons may gain energy equal to half the original energy of the ions, that is, 2 or 3 Kev.

Ferraro, V. C. A., On the theory of the first phase of a geomagnetic storm: A new illustrative calculation based on a idealised (plane not cylindrical) model field distribution, *J. Geophys. Res.*, 57, 15-49, 1952.

(Manuscript received June 6, 1959; revised November 13, 1959.)

Photographs of the High-Altitude Nuclear Explosion 'Teak'

W. R. STEIGER² AND S. MATSUSHITA

*High Altitude Observatory
University of Colorado*

Abstract. A sequence of four photographs of the August 1, 1958, high-altitude nuclear explosion 'Teak' near Johnston Island in the Pacific is shown. These photographs were taken from Maui, Hawaii, at an altitude of 3050 meters and 1300 km from the explosion. Three main features of these photographs are evident and are discussed: an auroral arc directed southward; an expanding envelope; and an airglow cloud. The auroral arc extending southward from the explosion is interpreted as the glow produced by a stream of β -decay electrons directed along the earth's magnetic field. This arc apparently extended into the southern hemisphere and was observed from Apia, Samoa. The expanding envelope is interpreted as an excitation-recombination phenomenon produced by an expanding shock front. By assuming an average speed of propagation of 1.3 km/sec of the shock front, the times at which the photographs were taken are estimated. The airglow cloud is interpreted as a residue of ionized material having a lifetime estimated at 15 to 30 minutes. Assuming an electron density of 10 times the normal F_2 -region value, an effective recombination coefficient of 10 times the normal F_2 value is obtained. The airglow cloud ascended at a rate of approximately 1000 m/sec and expanded horizontally at a rate of approximately 300 m/sec.

Accompanying this communication is a series of photographs, Figures 1 to 4, of the high-altitude nuclear explosion 'Teak,' which took place near Johnston Island, whose geographical coordinates are 16.7°N, 169.4°W, in the Pacific on August 1, 1958, at 10:50:05 UT [AEC reports B.39 and B.94, 1959]. These photographs were taken from an altitude of 3050 meters atop Haleakala on the Island of Maui, Hawaii, at the site of the Hawaii Institute of Geophysics and Seismology, Haleakala Observatory, 20.7°N, 156.3°W. The photographer, Mr. Walter Lang, used a Roliflex camera with a 75-mm, $f/3.5$ lens, making 10-second exposures on Royal-X Pan. The horizon in each photograph is a sea of clouds at an elevation of approximately 2400 meters. The camera was inadvertently moved slightly during the first and third exposures, as evidenced by star trails.

This explosion, together with the 'Orange' explosion of August 12 at the same location and the 'Argus' series in the South Atlantic, has been the subject of great interest among geophysicists because of the striking geophysical phenomena induced by the explosions, particu-

larly auroras, geomagnetic storms, and radio fadeouts. These phenomena for the Teak and Orange explosions are summarized and analyzed in a recent paper by Matsushita [1959].

The Atomic Energy Commission has reported in the press that the Teak explosion shown here took place at a distance of 800 miles (1300 km) from Hawaii. On the basis of the geomagnetic variations, Matsushita estimated the height of the explosion to be 70 to 80 km. This is well below the horizon seen in these photographs, which is about 150 km above sea level at the distance of the explosion.

Unfortunately, the exact times of these photographs are uncertain. It is known that the first photograph was taken shortly after the initial blast, and from evidence below we deduce that the others were taken at intervals of perhaps 1 or 2 minutes. On the basis of associated geomagnetic and ionospheric phenomena, the times of the photographs will be inferred later in this paper.

We shall discuss three main features of these photographs: (1) the auroral arc directed to the south in Figure 1; (2) the expanding envelope seen in Figures 2, 3, and 4; and (3) the airglow cloud seen in the center of each photograph.

¹Contribution 16, Hawaii Institute of Geophysics, University of Hawaii.
²On leave from the University of Hawaii.

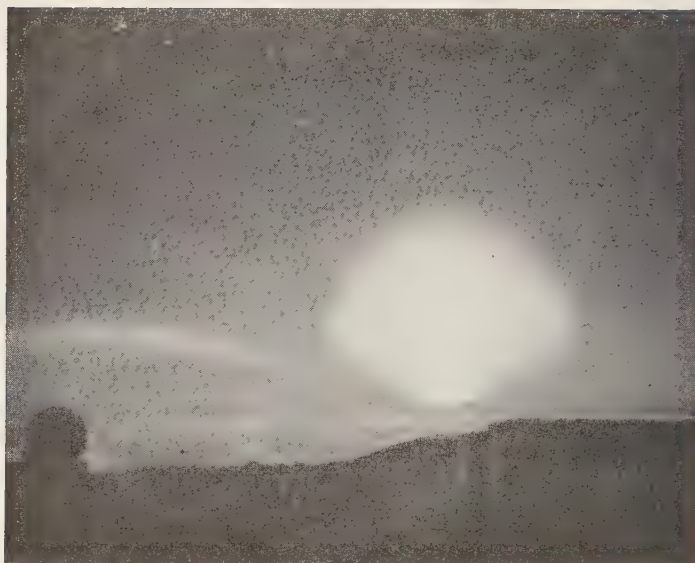


Fig. 1.



Fig. 2.

The auroral arc. The most striking feature of these photographs is the arc directed to the south in Figure 1. Without doubt this is an auroral glow produced by β -decay electrons ejected from the nuclear explosion directed

along the lines of the earth's magnetic field, postulated by *Fowler and Waddington* [1958] and *Kellog and others* [1959]. It is apparent the southern extremities of this arc which were reported by *Cullington* [1958] from Apia.



Fig. 3.



Fig. 4

Figures 1 through 4 are photographs of the nuclear explosion Teak as seen from Hawaii. Time sequence Figure 1 to 4. See text.

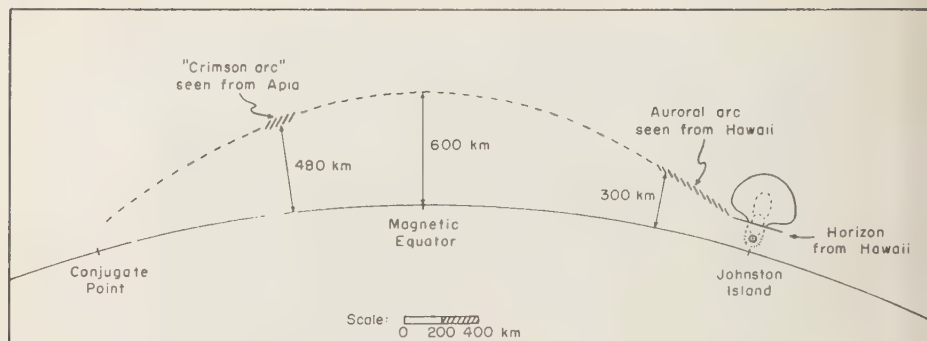


Fig. 5. The auroral arc.

The auroral arc and mushroom cloud of Figure 1 are drawn to scale in Figure 5, and the arc is extended to the conjugate point in the southern hemisphere. The photograph has been projected onto the plane of the magnetic meridian. The 'crimson arc' reported from Apia falls nicely onto this arc when using the position as determined by Matsushita. The height of the arc at the magnetic equator, about 600 km, agrees well with the height determined by Kellogg and co-workers. Because the Apia aurora seems to be associated with this arc, we assume that the photograph was taken at about the time the Apia aurora was observed, 10:51 UT.

The arc clearly originates from below the horizon in Figure 1, and its point of origin is

quite consistent with Matsushita's estimate of 70 to 80 km above sea level.

The expanding envelope. The expanding envelope seen in Figures 2, 3, and 4 is presumably an excitation-recombination phenomenon produced by the advancing front of a shock wave. If so, the average speed of propagation can be taken as about 1.3 km/sec, as calculated by Matsushita on the basis of the geomagnetic and ionospheric effects observed at Maui at the time of the passage of the shock wave. Since Figure 3 shows the shock front expanded to a diameter of about 900 km, this would place it at about 5.75 minutes after the explosion, or at 10:55 UT. Any time interval much longer than this would be clearly evident from the motion of the

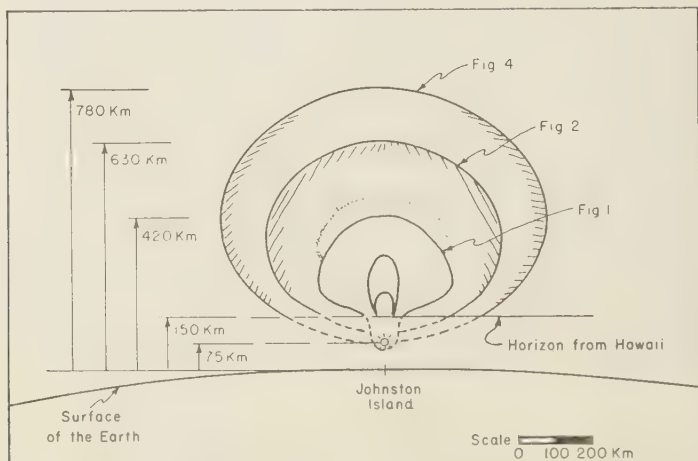


Fig. 6. The expanding shock front.

er field with respect to the horizon, but in the star field shows very little motion during the series. (The right-hand edge of Fig. 1 is almost due west.) By the same reasoning the times of Figures 2 and 3 can also be determined, being 10:54:10 and 10:55:30 UT, respectively. The mushroom cloud of Figure 1 is a transient phenomenon and does not appear in the other photographs. It is probably caused by the immediate emission of ionized particles or X-rays, or both, from the explosion and disappears as soon as the source of ionization dies out. On the original photograph a suggestion of the shock front is seen slightly above the mushroom cloud and of about the same diameter. If this is usually the shock front at this instant, it places the time of the photograph at 10:52:20. To summarize, the times we have estimated are:

Figure 1 10:52:20 UT
 Figure 2 10:54:10 UT
 Figure 3 10:55:30 UT
 Figure 4 10:55:50 UT

mentioned before, the estimated speed of 1.3 km/sec must be an average value only. The actual speed must have been somewhat greater initially, and somewhat less than this when passing over Hawaii. Thus, to the extent that the speed differed from the average value in the various photographs, the times given are somewhat in error. For example, if we assume that the initial speed was 2.0 km/sec and decreased linearly with time resulting in an average speed to Hawaii of 1.3 km/sec, the following times of the photographs result:

Figure 1 10:51:40 UT
 Figure 2 10:53:00 UT
 Figure 3 10:54:00 UT
 Figure 4 10:54:20 UT

The time obtained for Figure 1 by this approach is more nearly in agreement with the estimate from the auroral arc above. In Figure 6 we have drawn the mushroom cloud and expanding envelope to scale by projecting the photographs onto a vertical plane. The field of stars was used to determine the scale of the photographs. The shock-wave envelope has somewhat the form of an ellipsoid of

revolution with the major axis horizontal. The intensity of radiation is greatest about a horizontal ring at the major axis, 300 to 400 km above sea level. Near the center of the photographs we are looking directly through this ring so that it does not seem very bright, but at the edges of the ring it appears bright because we are looking through a much greater depth of the radiating gas.

From this composite drawing, the shock wave seems to have originated in the *F* region at a height of about 400 km. However, the photographs owe their appearance not only to the nature of the propagation of the shock wave but also to the nature of the ionosphere. Hence, what we see may be misleading, and in actuality the shock wave may have originated at a considerably lower level.

The airglow cloud. The persistent glowing central cloud that remains in Figures 2, 3, and 4 is very likely an artificial airglow resulting from a residue of ionized particles which recombine rather slowly owing to the low pressure at this altitude, which ranges from about 250 to 300 km. On the basis of a crude photometry of the negatives and reports from observers it appears that the lifetime of the cloud may have been of the order of 15 minutes and not more than 30 minutes. If we assume that the ionization density in the cloud was initially some 10 times the normal F_2 value of about 10^6 cm^{-3} , we must expect an effective recombination coefficient of approximately $10^{-9} \text{ cm}^3/\text{sec}$. This is about a factor of 10 greater than the normal F_2 value of the effective recombination coefficient, and may be accounted for by the heating and expansion of the cloud.

From Figure 7 we see that the airglow cloud

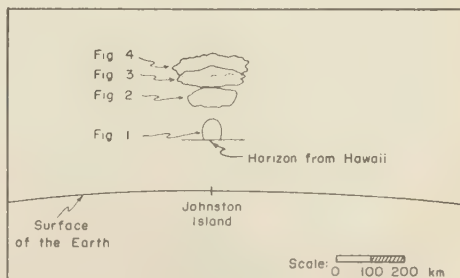


Fig. 7. The airglow cloud.

both rose and expanded horizontally. We can deduce the speeds of vertical rise and horizontal expansion by making use of the time intervals found above. We obtain a value of about 300 m/sec as the speed of horizontal expansion. This expansion must represent a radial wind caused by the rising and expanding gases from the explosion. The speed of vertical rise is found to be approximately 1000 m/sec.

From this necessarily crude analysis of these photographs it is strikingly clear how valuable would be a rapid sequence series of photographs taken at precisely known times, with photometric standards. The morphology and time sequence of the various events could be determined, the speed of the shock wave as a function of distance from the origin and height in the ionosphere, and perhaps something about the height and mechanism by which the shock

wave is generated, as well as other possible suits.

Acknowledgments. The authors are much indebted to Mr. Walter Lang for the use of photographs. The research reported in this paper has been sponsored by the National Science Foundation.

REFERENCES

- Cullington, A. L., A man-made or artificial aurora, *Nature*, 182, 1365-1366, 1958.
Fowler, P. H., and C. J. Waddington, An artificial aurora, *Nature*, 182, 1728, 1958.
Kellogg, P. J., E. P. Ney, and J. R. Winckler, Cosmic physical effects associated with high-altitude explosions, *Nature*, 183, 358-361, 1959.
Matsushita, S., On artificial geomagnetic ionospheric storms associated with high-altitude explosions, *J. Geophys. Research*, 64, 1149-1159, 1959.

(Manuscript received November 18, 1959.)

Balloon Observations of X-Rays in the Auroral Zone I¹

KINSEY A. ANDERSON²

*Physics Department,
State University of Iowa
Iowa City, Iowa*

Abstract. During high-altitude balloon flights made in the northern auroral zone, fluxes of X rays have frequently been observed. By means of sodium iodide scintillation crystals and pulse height analyzers also carried in the balloon instrumentation detailed measurements of the photon energy spectrum in the region 40 to 340 kev have been obtained. These X rays generally appear without definite geophysical or solar correlation. They are believed to be directly related to the soft radiation investigated by rockoons and also to processes occurring in the outer Van Allen radiation zone. The rapid intensity and spectral fluctuations that are observed suggest that in addition some process, possibly atmospheric discharge, may also be important. The relation of these X rays to visible auroras is not clearly established here.

INTRODUCTION

During the past two years thirty-seven constant level balloon flights of long duration have been made from sites in the northern auroral zone. Each of these flights carried at least two, and most of them three, charged particle detectors. During many of these flights sporadically appearing fluxes of X rays have been encountered. They are usually characterized by large and rapid intensity fluctuations during the time of their appearance, which is typically several hours. It is believed that the X-ray effects are produced by electron *bremsstrahlung* at higher levels in the earth's atmosphere. One of these events occurred in close association with the geomagnetic storm of August 29-30, 1957, and has been described previously [Anderson, 1958]. The five additional occurrences of X rays in the auroral zone which will be described here have neither very definite correlations with other geophysical phenomena nor obvious solar association. This is in marked contrast to the very intense X-ray bursts described by *Winckler and others* [1958, 1959] that appear in connection with the spatially large and active auro-

ral displays that occur during world-wide magnetic disturbances. The connection of the X rays observed in the high-latitude measurements with visible auroras is not established here, but they are believed to be directly related to the soft radiation investigated by rockoons [Meredith and others, 1955; Van Allen, 1957] and to processes occurring in the outer Van Allen radiation zone [Van Allen and Frank, 1959].

RESULTS

Flight 63, September 1-2, 1957, Fort Churchill, Canada. In Figure 1 is seen the counting rates of the three charged particle counters carried on that flight plotted against universal time. The absence of effects in the counter telescope shows that only a small part of the effect in the single counter or in the ion chamber can be due to charged particles. After about 1600 UT the X-ray effects are seen to decrease gradually but with sharp fluctuations superimposed on this general trend. Since the balloon is slowly increasing its depth (about 0.2 g cm⁻²/hr) some of the decrease in X-ray intensity may be due to this.

There was no definite geomagnetic disturbance that could be associated with the appearance of these X rays, although the period from August 28 into the first week of September saw the occurrence of very great solar activity and

¹This work was supported by the Office of Naval Research and its Skyhook program and is part of the United States participation in International Geophysical Cooperation 1959.

²Now at Department of Electronics, Royal Institute of Technology, Stockholm 70, Sweden.

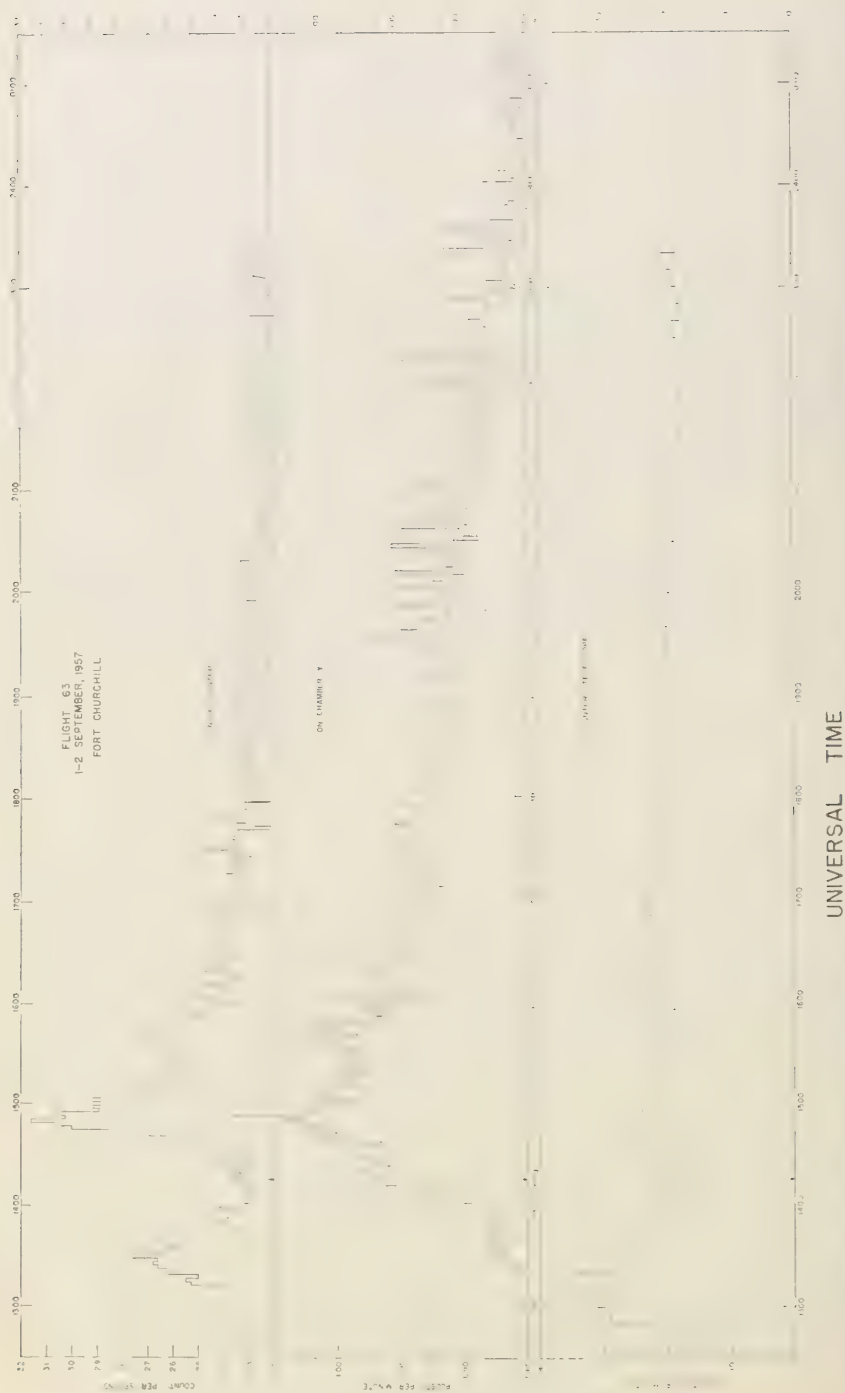


Fig. 1. X-ray and possibly some solar proton effects during a flight made at Fort Churchill. The very sharp spikes seen in the ion chamber are due to rapid and short-lived increases in the X-ray flux. These spikes are superimposed on a gradual decline of the X-ray flux which lasts many hours. The base lines of the single counter and ion chamber plot represent approximately the normal cosmic-ray levels.

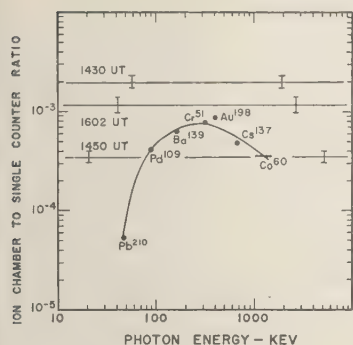


Fig. 2. Calibration of the detector used on flight 63 with X-ray sources. The horizontal lines indicate various values of the ion chamber to single counter ratio encountered during that flight.

associated terrestrial effects. During most of this period cosmic noise absorption measurements at 30 Mc/s showed deep and long-enduring attenuation [Hartz, private communication]. This absorption was probably due to an influx of solar protons [Anderson and others, 1959] of energy too low to reach the detectors carried on this flight. At the depth of the balloon this would correspond to an energy of less than 110 Mev above the atmosphere. Considerable auroral activity was present during the flight of August 31–September 1, and large magnetic bays were observed up to 0900 UT on September 1.

The ratio of the ion chamber pulse rate to the single counter counting rate is a function of the charge as well as of the energy of the particles entering it. This ratio also takes on definite values for photons, as can be seen in Figure 2. Here calibration of an ion chamber and single counter has been made with a variety of gamma-emitting sources. The ion chamber to single counter ratio for the counting rate effects above the cosmic-ray background is shown in Figure 3. It is seen from this figure and from Figure 2 that during the large peak at 1450 UT the response is characteristic of 80-keV X rays. The curve is double valued, but the lower energy is preferred mainly on the basis of subsequent experiments. It will also be noticed that from about 1410 to 1435 UT the ratio is too large to be attributed to photons. Again, past 1600 UT, the ratio becomes quite high. A large ion chamber to single counter ratio, that is,

considerable ionization from relatively few particles, is characteristic of protons or heavier nuclei. The counter telescope data would ordinarily be conclusive in this matter, but its geometry factor is too small to obtain adequate statistical accuracy for choosing between X rays and charged particles. The best interpretation that can be given this event at present is that the counter effects are mainly due to X rays but at times, notably at 1430 and possibly around 1600 UT, solar protons, associated with the polar cap (type III) cosmic noise absorption [Reid and Collins, 1959; Leinbach and Reid, 1959], reach the detectors. Several balloon flights [Anderson and others, 1959; Brown and D'Arcy, 1959; Anderson and Enemark, unpublished] have shown that, when polar cap cosmic noise absorption of several decibels is present, high fluxes of protons above 100 Mev kinetic energy before entering the atmosphere are also present. During flight 63 when radio noise absorption of 4 db magnitude at 30 Mc/s was recorded above Fort Churchill [Hartz, private communication], no proton fluxes were observed except for the brief times noted above. This result could be explained if the energy spectrum were somewhat steeper on the occasion of flight 63 than on the later flights referenced above, which give a power law exponent for the differential energy spectrum of about 5.

It is not possible at present to decide whether the X rays during flight 63 are associated with the supposed influx of solar protons or with the soft radiation phenomenon. In the light of later experience with both solar proton events and auroral zone X-ray occurrences the latter possibility seems the more probable.

Flight 978, April 14–15, 1959, Fairbanks, Alaska. The measurements shown in Figure 4 were made with a sodium iodide scintillation crystal, which is approximately 200 times more efficient to 100-keV X rays than the ion chamber. The associated transistorized circuitry has been described by Enemark [1959]. The output pulses from the photomultiplier were sorted into four integral pulse height channels whose edges are given in Figure 4. The entire balloon flight is shown there so that the smooth peak on the left whose maximum is seen to occur in all channels at about 1830 UT is the high-energy cosmic-ray transition effect (Pfotzer maximum) encountered during the ascent. The X-ray

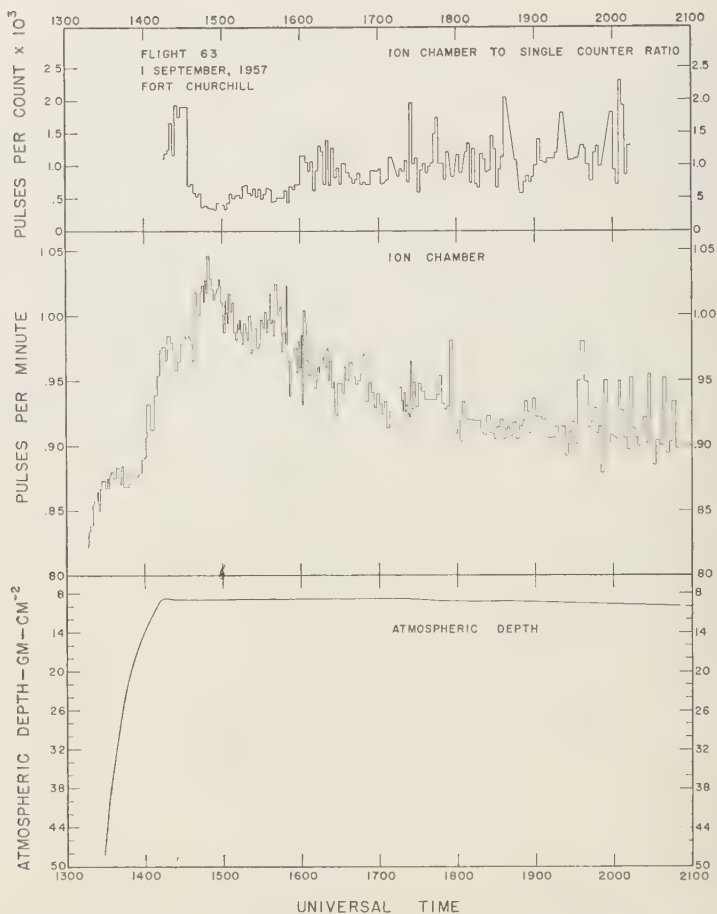


Fig. 3. Large variations are seen in the ion chamber to single counter ratio during flight 63. Past 1700 UT the effects in the single counter become quite small, with the result that large statistical errors occur in the ratio.

effects are seen to begin at about 0010 UT and continue until the flight ends just past 1140 UT, a total of $11\frac{1}{2}$ hours.

Absence of effects in coincidence counters and gaseous detectors leads to the conclusion that the effects above the cosmic-ray background must be due entirely to X rays. The two features of greatest interest in this event are the very sharp fluctuations in the X-ray intensity and the changes in their energy. The spike that occurs at 0143 UT is plotted in Figure 5 with the greatest available time resolution. It is seen to be about 30 seconds in width halfway from the root to the peak. The topmost part of the

peak is 2.5 seconds in width. This interval is the time between successive outputs from a scale 512. Figure 5A shows the photon effects in the higher energy channels. To produce an X-ray pulse of 30 seconds half-width at the position of the balloon, a source of electrons at a height of 100 km would need to drift with a velocity of 6 km/sec. However, the shape of the expected X-ray burst calculated on this basis does not agree very well with the observed peak shown in Figure 5.

During the sharp spike occurring at 0143 UT there is a small effect in the highest energy channel, showing that electrons of at least 3

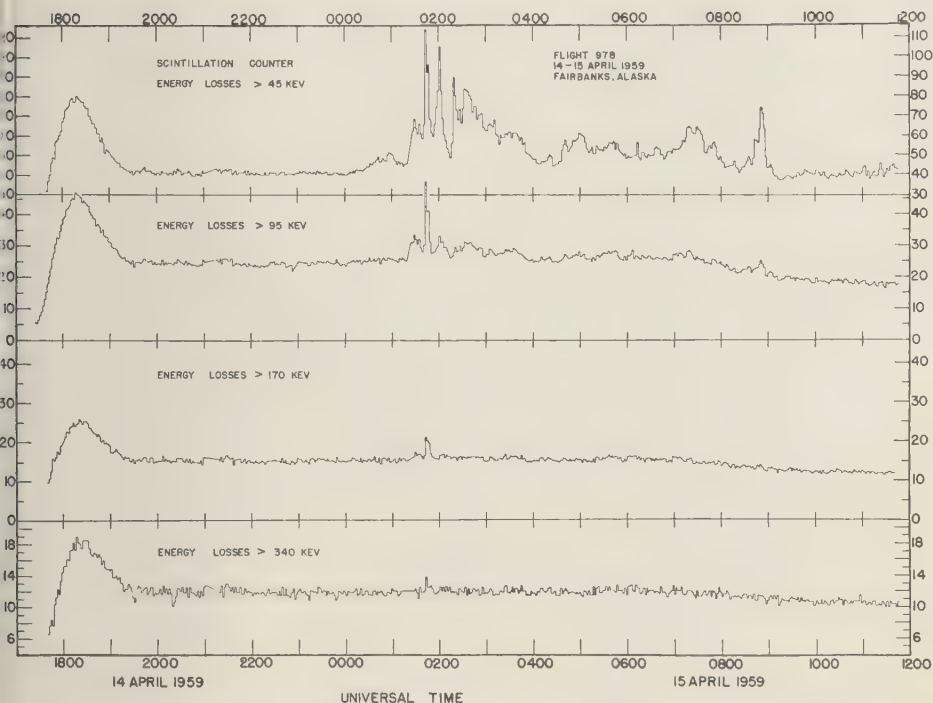


Fig. 4. Integral counting rates from the scintillation counter and four-channel pulse height analyzer. The decrease in the counting rate due to normal cosmic rays past about 0700 is accounted for the very large temperature drop (about 60°C) in the instrument package that occurs after sunset.

v and very probably of much higher energy e involved above the atmosphere. It is unlikely that the effect in this channel is due to le-up from a large number of small pulses, en below the lowest discriminator edge, since rger X-ray effects above 40 kev have been en on other flights but with no effects present the highest energy channel.

The large changes occurring in the photon ergy spectrum at the time of this peak are st seen in Figure 6. Here both the integral otion flux and the differential photon intensity ve been plotted for three different times during the event. The integral flux is defined as lows:

$$N(> E_i) = \frac{T(E_i) - C(E_i)}{\epsilon_i G}$$

E_i) is the whole counting rate in counts per second of the crystal due to energy losses greater than energy E_i kev.

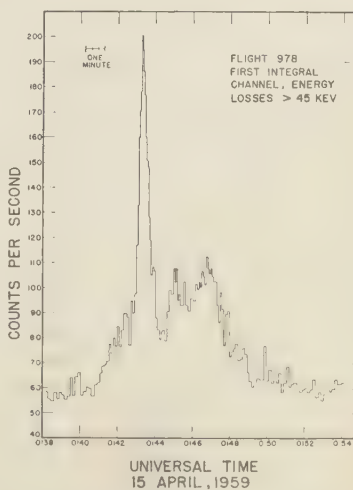


Fig. 5. An X-ray burst plotted with the maximum available time resolution. During this spike the photon energy spectrum changes markedly.

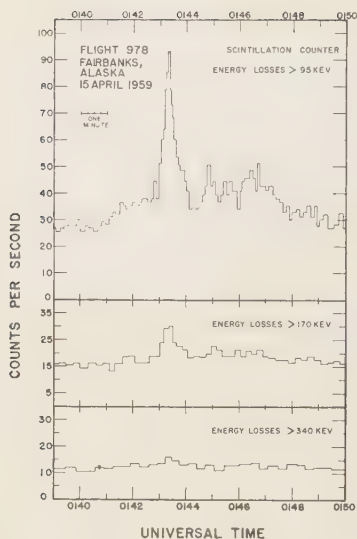


Fig. 5A. The same X-ray burst shown in Figure 5 as seen in the higher energy channels of the pulse height analyzer.

$C(E_i)$ is the counting rate due to energy losses greater than E_i kev contributed by the high-energy cosmic-ray background.

ϵ_i is the efficiency for converting photons having energy above E_i to electrons in the crystal. In principle the spectrum must be known and the weighted efficiency found, but in practice the spectrum is so steep and the efficiency versus photon energy such that it is near unity for all integral channels, except the highest energy, where it is about 0.45.

G is the area of the crystal weighted isotropically with respect to the solid angle. It can be determined from the formula

$$G = 1/4\pi al(l + a/21)$$

where a is the crystal diameter and l its height. For the $3/4$ by $3/4$ inch crystal employed, G is 4.25 cm^2 .

The found differential photon intensity is then found as follows:

$$n(E_{i+1}, E_i) = \frac{N(> E_i) - N(> E_{i+1})}{E_{i+1} - E_i}$$

$$i = 1, 2, 3$$

Figure 6 then shows the photon spectra at

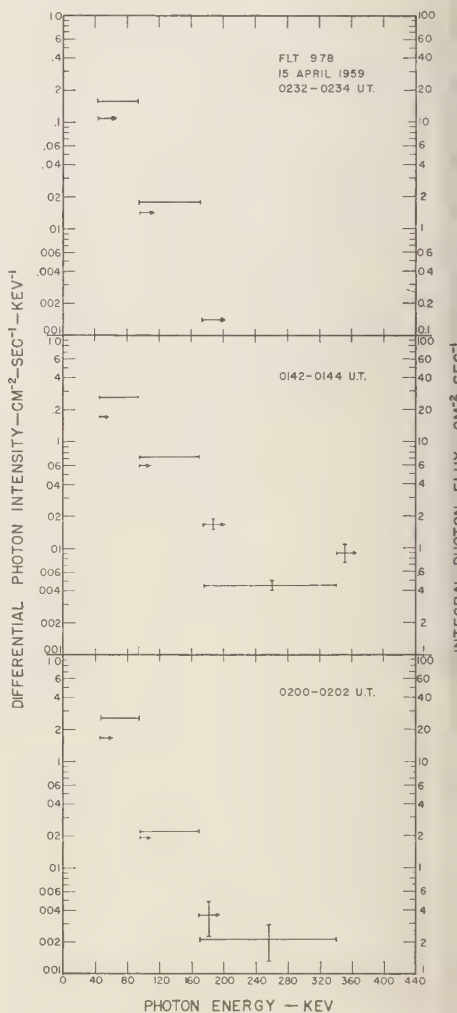


Fig. 6. Photon energy spectra at three different times during flight 978 made in Alaska. The 0142-0144 UT spectral determination corresponds to the spike shown in Figure 5.

particular times during this event. They are obtained by averaging the various channel counting rates over 2-minute intervals. From 0232 to 0234 UT the spectrum is typical of most of the flight, but during the appearance of the X-ray spike from 0142 to 0144 UT there is a remarkable spectral change. A strong flattening occurs in the first three integral channels

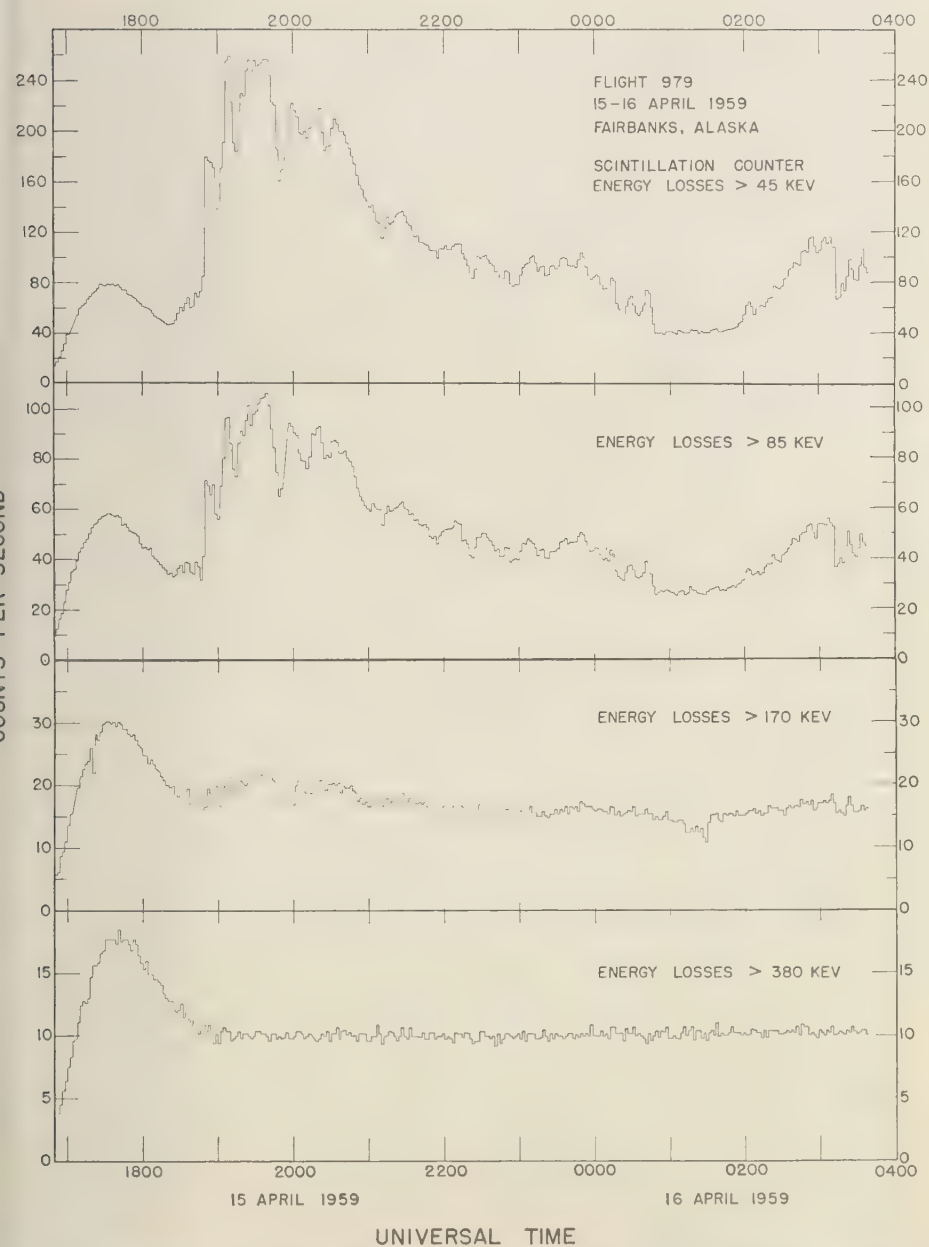


Fig. 7. X-ray effects during Alaska flight 979 launched 5 hours after the termination of flight 978 shown in Figure 4.

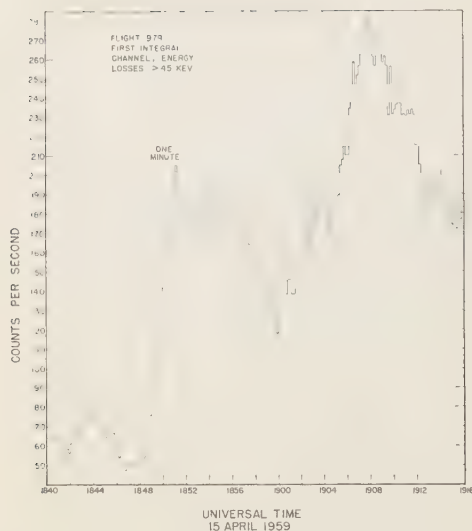


Fig. 8. The time structure of the X rays during flight 979 is not as rapid as that observed during flight 978.

and also a 'toe' containing photons of above 340-keV energy appears in the highest energy channel. The peak occurring at 0201 UT exhibits a very steep spectrum compared even with the 0233 UT determination.

The atmospheric depth maintained by the balloon during these measurements was between 8.7 and 9.1 g cm⁻², but no attempt has been made to correct the spectra to smaller depths since this would involve assumptions about the angular distribution of the X rays. Therefore all the photon energy spectra presented here refer to the atmospheric depth of the balloon, and in addition about 0.4 g cm⁻² of material equivalent to aluminum surrounds the sodium iodide crystal. The effect of this matter is to reduce the number of photons reaching the sodium iodide crystal. And since the absorption coefficients are greater at the lower X-ray energies the observed spectrum will be a flattened version of the original spectrum, the flattening increasing as the photon energy is lowered.

Flight 979, April 15-16, 1959, Fairbanks, Alaska. The counting rates in the four integral channels for this flight which reaches ceiling 7 hours after the termination of the preceding flight are shown in Figure 7. The X-ray effects

appear even before the balloon has reached ceiling altitude, which for this flight is only about 3 g cm⁻². The balloon rose slowly during the day, and at the end of the flight the depth was 2.7 g cm⁻². It will be noticed that at no time during this flight are effects in the highest energy channel observable. This fact, in addition to the result that the effect in the ion chamber can be quantitatively attributed to the X rays, permits a limit being set on the flux of electrons that have sufficient energy to penetrate the residual atmosphere and the walls of the apparatus without undergoing *bremsstrahlung*. Since this flight is a very high one and the X-ray fluxes large, the limit established is about as good as can be obtained now by the balloon technique. The absolute flux of electrons having energy above 7 Mev outside the atmosphere that would readily be observed during the X-ray peak occurring at 1908 UT on flight 979 is somewhat less than 0.1 cm⁻² sec⁻¹. The X-ray flux above 45-keV photon energy at this time is 65 cm⁻² sec⁻¹. Thus there is less than 1 electron above 7 Mev outside the atmosphere for each 650 photons above 45 keV at a depth of 3 g cm⁻².

The fluctuations in the X-ray intensity, though still rapid, are by no means as striking as in the previous flight (978). The structure of some of the X-ray peaks is plotted in Figure 8. It is seen that the 'rise time' is about 90 seconds compared with 20 seconds for the large spike during flight 978 (Fig. 5). A small effect is produced in the ion chamber by these X rays, as can be seen in Figure 9. Also shown there is the ratio of the integral X-ray flux in the first channel to that in the second. This of course does not represent all the information available about the time changes in the photon energy spectrum but does indicate that the spectrum, although slowly varying, does undergo time changes. The spectrum has been determined in detail at three different times during flight 979; these results are presented in Figure 10. It is seen there that for these times the spectrum shows much the same shape.

Associated with the X-ray fluxes observed on this flight there is a definite cosmic noise absorption at 27.6 Mc/s amounting to a maximum of 1.2 db at 1930 UT [Leinbach, private communication]. Furthermore, the profiles of the X-ray counting rate and the radio noise ab

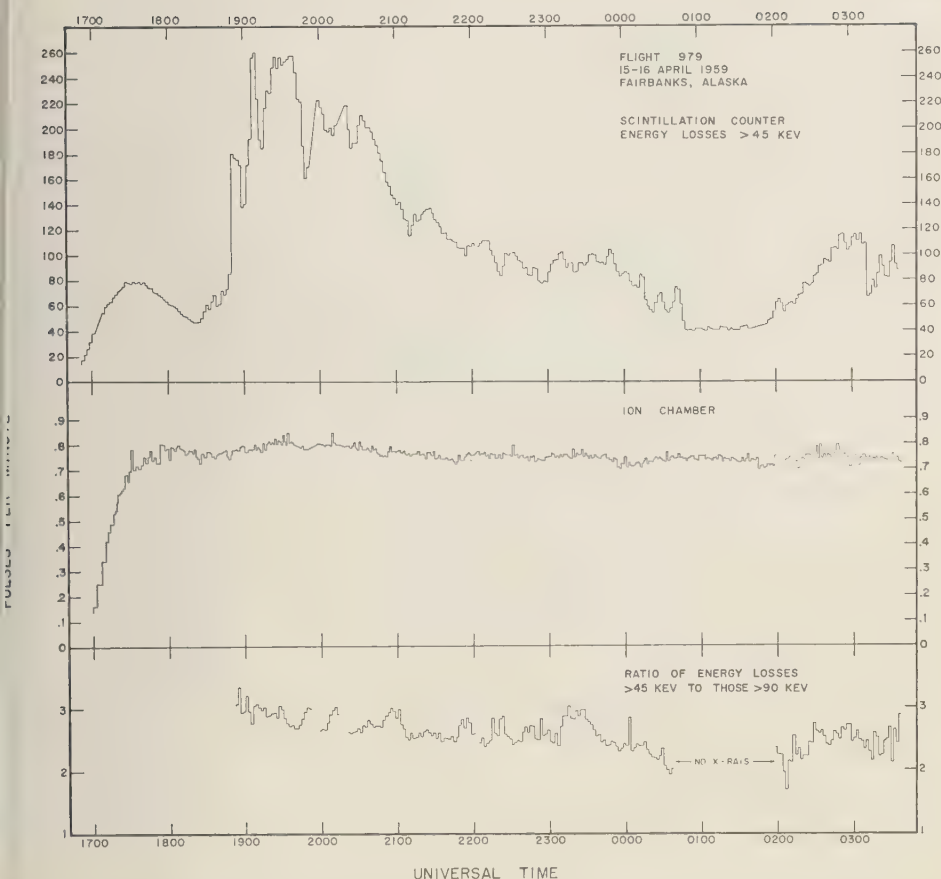


Fig. 9. This figure shows that the X-ray effects in gaseous detectors are only a few per cent whereas in the crystal scintillator they are several hundred per cent. In the bottom diagram it is seen that the energy spectrum undergoes significant changes with time.

absorption are quite similar over a several-hour period. Because of the limited sensitivity of spectrometer devices, however, no detailed comparison can be made.

Flights 980 and 982, Fairbanks, Alaska. Each flight shows the appearance of relatively moderate X-ray effects lasting for approximately 3 hours as is shown in Figures 11 and 12. The general outlines of these X-ray influxes are similar to but much smoother than those exhibited by the two previous Alaska flights (978 and 979). Because of low temperatures encountered during the ascent part of flight 980, one of the binary scaling circuits was inoperative until

about 1700 UT. This accounts for the abnormal appearance of the 40-kev channel in Figure 11, and in particular for the absence of the Pfotzer maximum. Past 1700 UT there is every indication that the circuits were performing normally. The integral photon fluxes and differential photon intensities have been plotted versus photon energy in Figure 13. The X rays appeared at a time during flight 980 when the depth of the balloon was 2.5 g cm^{-2} , and during flight 982 at a depth of 8.9 g cm^{-2} .

These spectra seem to be somewhat flatter than those obtained for flight 979 and shown in Figure 10, although still within the spectrum

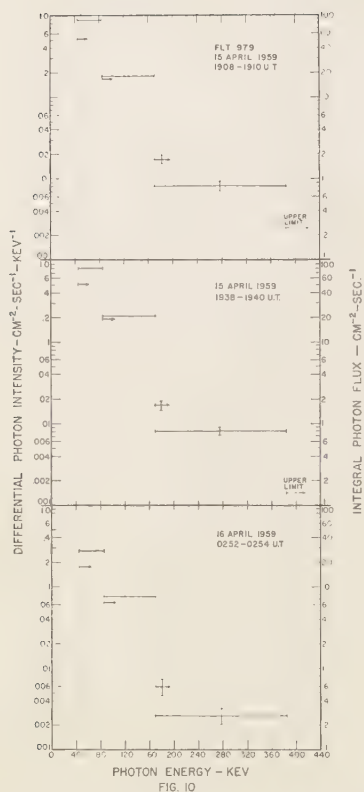


Fig. 10. Photon energy spectra obtained during flight 979. The spectral changes over this flight are not as large as on flight 978.

fluctuations as judged from the channel I to channel II ratio shown in Figure 9.

DISCUSSION

1. From 70 hours of balloon data obtained at altitudes greater than 105,000 feet above Fairbanks, Alaska, during mid-April, 1959, sizable X-ray fluxes of a sporadic nature were present for 25 hours, or 35 per cent of the total time monitored. The X-ray effects begin in all examples presented here during the daytime, although the balloon flights were heavily biased in this direction since of the 70 total hours of high-altitude data 55 hours were obtained while the atmosphere above the balloon was sunlit. It is worth noting that the X-ray fluxes during flight 978 are smaller after the sun has set than

during the daytime. This of course is not necessarily a genuine day-night effect but taken in connection with several other balloon flights is an example of the empirical rule that the biggest X-ray effects are observed in the daytime. Additional data bearing on this point will be presented in the second part of this paper, to be published later.

These auroral zone X-ray fluxes are typically present for a few hours each day and therefore contrast with the magnetic-storm-correlated auroras that appear farther south, which yield X rays on the order of a few hours each month [Winckler and others, 1958, 1959].

2. These observations of X-ray fluxes were made at a time when visible auroral displays were few and quite inactive. This period was also one of little geomagnetic activity. Furthermore, the X-ray fluxes, particularly the large bursts, had no connection with the small disturbances present on the magnetometer traces during the time of the flights. This is in marked contrast to the magnetic storm X rays of August 29, 1957 [Anderson, 1958], and February 10-11, 1958 [Winckler and others, 1958]. Several days before the Alaska balloon flights there was a period of very considerable geomagnetic activity during which there were frequent appearances of bright and active visible auroras. The A_p values for the period April 9 to 20 and the K_p values during the time of the balloon flights are given in Table 1. For flight

TABLE 1*

April	7	8	9	10	11	12	13	14
A_p	8	22	44	98	23	10	9	11
K_p Values								
Three-hour Greenwich Interval								
April	1	2	3	4	5	6	7	8
15	4	2	2	1	1	2	3	2
16	2	1	2	3	1	1	2	1
17	0	1	3	3	3	1	2	2
18	2	1	1	1	1	2	2	0
19	1	1	1	1	1	2	1	2
20	1	1	1	2	2	1	1	1

* These data obtained from National Bureau of Standards report CRPL-F181 Part B *Solar-Geophysical Data* issued September 1959.

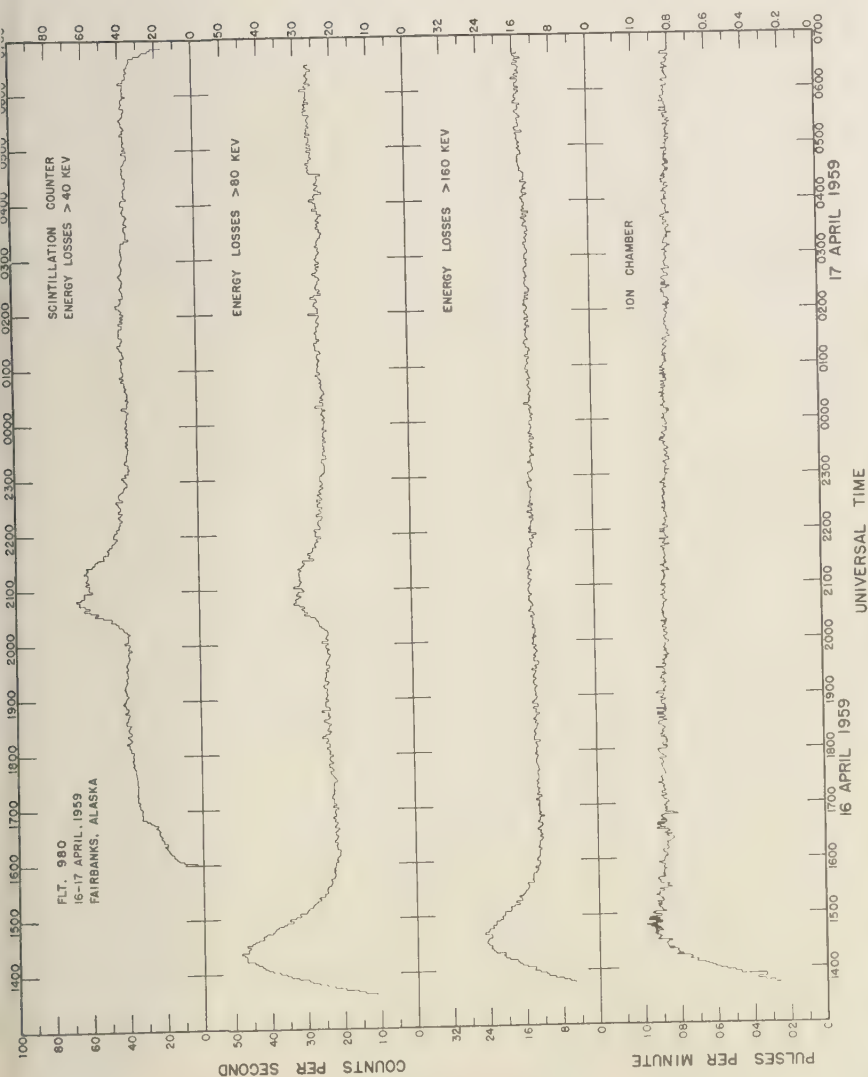


Fig. 11. The influx of X rays during this flight is seen to be comparatively smooth. The abnormal appearance of the first integral channel is due to the very low temperatures encountered.

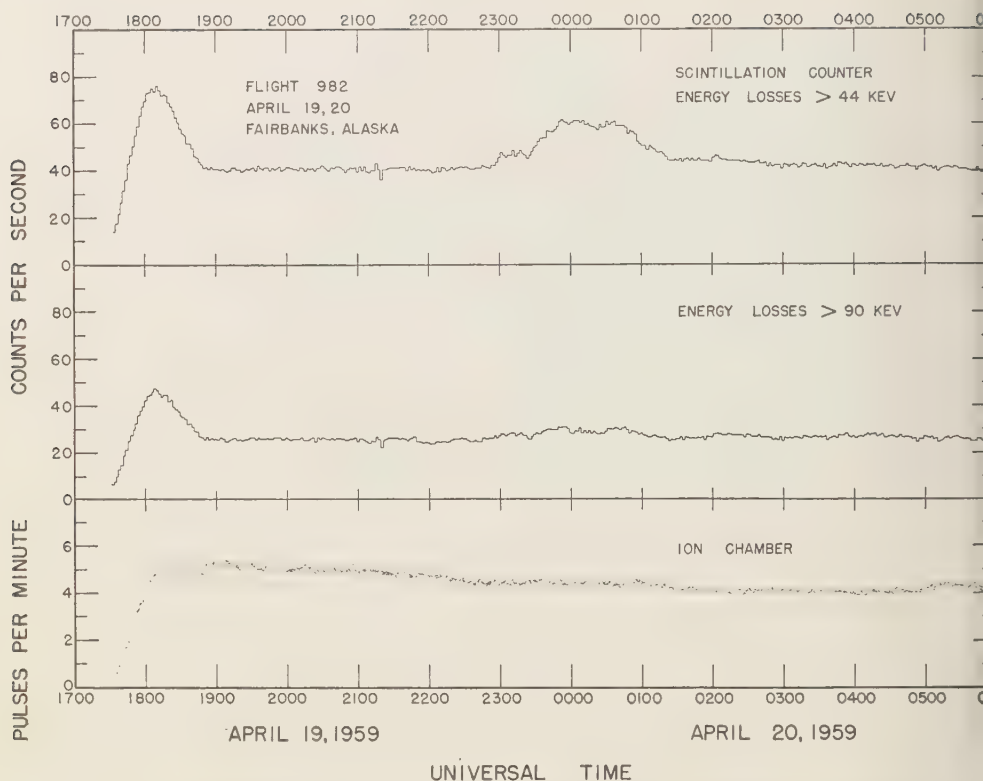


Fig. 12. The appearance of the X rays on this flight is much the same as during flight 980.

63 made on September 1, 1957 (Figs. 1 and 3), the K_p value averaged 2 during the time X rays were present.

3. The photon intensities observed on the four Alaska flights reported here are much smaller than those that occurred over the auroral zone during the magnetic storm of August 29, 1957, and beneath the active auroral displays that appear far to the south of the auroral zone. Most of the effects observed are undetectable in the argon-filled ion chamber, but the largest X-ray bursts produce about a 10 per cent increase in the pulsing rate above the cosmic-ray background. The effects observed during the magnetic storm on August 29 above Fort Churchill were up to 20 per cent in an ionization chamber, and auroras above Minneapolis give effects characteristically of this size and much larger. Increases of approximately 600 per cent were observed during the

storm of September 22–23, 1957 [Winckler and others, 1958], and 100 per cent effects appeared on February 10–11, 1958 [Winckler and others, 1959].

The auroral-zone effects reported here typically exhibit X-ray fluxes of 10 to 100 photons $\text{cm}^{-2} \text{sec}^{-1}$ above 40 keV as obtained from the scintillation counter.

4. The X-ray fluxes observed in the five flights presented here are probably to be associated with the soft radiation investigated by rockets. The soft radiation was observed to appear in Geiger counters at altitudes as low as about 40 km. With the far more efficient sodium iodide scintillation counters employed here it is entirely reasonable that effects should be seen at atmospheric depths up to 10 g cm^{-2} . Another strong indication that the X-ray showers are related to the soft radiation is that the geographical distribution is quite similar. The basi-

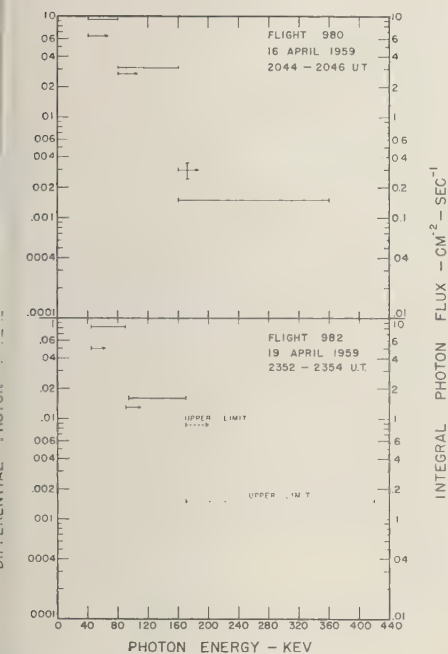


Fig. 13. The photon energy spectra for the smooth X-ray influxes on flights 980 and 982.

or this statement is that 160 hours of high-latitude balloon data, including 60 hours of observation with a scintillation counter, above Resolute Bay near the north magnetic pole detected no X-ray fluxes at any time during and following a period of considerable magnetic disturbance.

The Alaska results do not establish a direct connection between the X-ray bursts and visual auroras because of the fact that the flights were mainly in the daytime.

5. The auroral zone X-ray fluxes as observed in long-duration balloon flights often exhibit a small time scale. Superimposed on one several hours long are sharp fluctuations, which almost invariably begin as increases, that occur within a few very few minutes and often within seconds. Of the flights presented here, 63 and 978, shown in Figures 1 and 4 respectively, are the best examples of this effect. The rapid intensity variations are probably the most striking general feature of these observations, particularly since they occasionally involve large changes in

the photon energy spectrum, as during flight 978. Although no quantitative theory exists on the point it would appear difficult to account for this very rapid behavior on the basis of a simple 'dumping' of particles from the outer radiation zone. Explanation of rapid spectral changes ought to be especially intractable on such a view. The observed behavior of the X-ray fluxes might best be explained first by relating them to processes in the outer Van Allen radiation belt and secondly by invoking a local means of accelerating particles in or entering the radiation belt. As was mentioned above, balloon observations over the north magnetic pole showed that X rays did not appear there even during and following periods of geomagnetic disturbance, in contrast to the regular appearance of X rays in the auroral zone. This result evidently means that clouds of magnetized solar gas containing trapped high-energy electrons which are released upon the entry of the cloud into the earth's magnetic field are not the direct cause of the auroral-zone X-ray effects, assuming that such clouds could arrive as readily to the north of the auroral zone as within it.

As to a local process that might provide the rapid intensity and spectral changes exhibited by the data presented here no conclusive statement can be made, but the acceleration of electrons by means of an electric field in the atmosphere is an attractive possibility. Such a theory has been outlined by Reid [1958] to interpret auroral morphology. The attractiveness to the X-ray problem lies in the natural way by which rapid intensity and spectral fluctuations could be qualitatively explained. Reid's theory also predicts day-night differences in the auroral phenomenon, and, as is here suggested by flight 978, the balloon observations of X rays in the auroral zone indicate more intense X-ray effects during the day. Further data on this point will be presented in the second part of this paper.

Acknowledgment. The Skyhook field operations were conducted by General Mills, Incorporated. The support of the Geophysical Institute at College, Alaska, Dr. C. T. Elvey, director, was invaluable to the conduct of these experiments. The University of Iowa personnel who assisted in the Alaska experiments were Mr. Donald C. Enemark and Mr. Burrell E. Mohler.

REFERENCES

- Anderson, K. A., Soft radiation events at high altitude during the magnetic storm of August 29-30, 1957, *Phys. Rev.*, *111*, 1397-1405, 1958.
- Anderson, K. A., R. Arnoldy, R. Hoffman, L. Peterson, and J. R. Winckler, Observations of low-energy solar cosmic rays from the flare of 22 August 1958, *J. Geophys. Research*, *64*, 1133, 1959.
- Brown, R. R., and R. G. D'Arcy, Observation of solar flare radiation at high latitude during the period July 10-17, 1959, *Univ. Calif. Dept. Phys. Preprint*, September, 1959.
- Enemark, D., Balloon-borne circuits sort high-altitude cosmic rays, *Electronics*, *32* (35), 52, 1959.
- Leinbach, H., and G. C. Reid, Ionization of the upper atmosphere by cosmic rays of solar origin, *Phys. Rev. Letters*, *2*, 61, 1959.
- Meredith, L. H., M. B. Gottlieb, and J. A. Van Allen, Direct detection of soft radiation above 50 kilometers in the auroral zone, *Phys. Rev.*, *97*, 201, 1955.
- Reid, G. C., Electric field theory of aurorae, *Nature*, *182*, 1791, 1958.
- Reid, G. C., and C. Collins, Observations of abnormal VHF radio wave absorption at medium and high latitudes, *J. Atmospheric and Terrestrial Phys.*, *14*, 63, 1959.
- Van Allen, J. A., Direct detection of auroral radiation with rocket equipment, *Proc. Natl. Acad. Sci. U. S.*, *43*, 57, 1957.
- Van Allen, J. A., and L. A. Frank, Radiation measurements to 658,300 kilometers with Pioneer IV, *Nature*, *184*, 1959.
- Winckler, J. R., L. Peterson, R. Arnoldy, and R. Hoffman, X rays from visible aurorae at Minneapolis, *Phys. Rev.*, *110*, 1221, 1958.
- Winckler, J. R., L. Peterson, R. Hoffman, and R. Arnoldy, Auroral X rays, cosmic rays, and related phenomena during the storm of February 10-11, 1958, *J. Geophys. Research*, *64*, 597, 1959.

(Manuscript received October 29, 1959;
revised December 1, 1959.)

Some Spectral Studies of the Aurora¹

R. C. BLESS,² C. W. GARTLEIN, AND G. SPRAGUE

*Cornell University
Ithaca, New York*

Abstract. Some results of spectral studies of the aurora are discussed. The spectra, taken routinely since 1941, give a continuous spectral record of the north sky for every night. Special emphasis was given the study of hydrogen, with two main results: hydrogen appears in the early part of an aurora, and may be gone before the visual display is over; the absolute intensity of an aurora increases as the aurora expands outward from the pole. Sample intensity measurements are given showing the time variation of various components.

Introduction. Since 1941 continuous spectral records of the northern sky have been made at Ithaca, New York. Some of them, chosen on the basis of the presence of the H_{β} line and of auxiliary data, have been reduced to give information about the source intensity as a function of time and type of excitation. The specific features investigated included the relative behavior of hydrogen lines and the nitrogen bands, and the correlation of intensity with spectral appearance [Bless, 1954].

The spectrograms used were made with a $f(0.7)$ prism [Gartlein, 1944] spectrograph which records the auroral spectrum from 3914 Å to about 6800 Å on Eastman type I-C or 103aF photographic plates with an average dispersion of about 520 Å/mm. The plate moves downward automatically in the focal plane of the instrument (perpendicular to the dispersion) at the rate of $\frac{1}{2}$ inch per hour, or about one slit length in 10 minutes. Thus a continuous record of the auroral spectrum versus time is obtained. Exposures last all night. Impressed on each plate are a neon fiducial spectrum and a step wedge for intensity calibration.

The spectrograph was pointed at the northern sky 8.5° above the horizon so that the auroras seen by the instrument are characteristic of about 60° geomagnetic latitude (about 360–1100 m away). Figure 1 shows a microphotometer trace of a typical plate, for a given hour.

Reduction of data. In order to convert from the photographic density, measured by a micro-

photometer, to the actual source intensity, many factors must be considered. The C plate data were first converted to equivalent F plate data from the manufacturer's plate response curves. The conversion from density to intensity of plate illumination was accomplished by means of the calibrated step wedge on each plate. Since the step wedge was applied with green light, conversions were necessary when dealing with other spectral regions. These were taken from the spectral sensitivity curves of the F plate, as supplied by the manufacturer. The use of these curves implies the assumption that gamma of the emulsion does not vary seriously with wavelength. (The various photographic and photometric features, including reciprocity failure, limit the accuracy of the procedure, but their effects are probably not important in relation to other, unknown factors involving the aurora itself.)

When a value of the illumination at the plate has been obtained, it is necessary to determine the intensity at the spectrograph, or, in other words, to correct for the instrumental transmission. This conversion from plate intensity to slit intensity was made by determining the percentage transmission of the spectrograph as a function of wavelength. This was accomplished using radiation from a monochromator as the source and comparing the intensity at the focal plane with the slit intensity, both determined photoelectrically. Care was taken to maintain constant optical geometry in the measurements. This procedure gave the transmission curve of the spectrograph shown in Figure 2. Because of the glass lenses and prism, the instrumental

¹ This research was supported in part by the National Geographic Society.

² Now at University of Wisconsin.

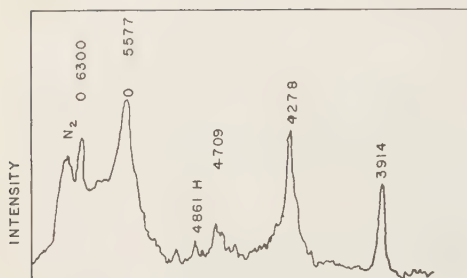


Fig. 1. Microphotometer trace of typical auroral spectrum.

transmittance is falling rapidly in the region of 4000 Å, which is unfortunate since the radiations in this region are of considerable interest.

To obtain the auroral intensity, knowing the intensity at the slit of the spectrograph, it was necessary to take the atmospheric absorption into account. This factor was important, since the zenith angle of the auroral forms seen by the instrument is large and the absorption depends strongly upon the zenith angle. The mechanics of the correction is briefly as follows. The orientation and angular width of the spectrographic cone of acceptance were determined by experiment, photoelectrically as before. This cone was defined as the region in which the focal plane intensity was more than one-half the maximum. It was found to be 8.5° in altitude and 6.0° in azimuth.

The atmospheric extinction was determined for the ray along the spectrograph axis and an aurora 100 km above the earth. This extinction for 8.5° elevation was taken from the tables of van de Hulst. As will be shown later this leads to error, since the absorption for an aurora off the axis but still in the cone of acceptance may differ markedly from the absorption for one on the axis. The intensity measurements are considered to have an over-all error of ± 25 per cent, excluding possible errors caused by the uncertainty in the zenith distance of the auroral form.

Results of analysis of plates. A sample of the analysis for a few nights is shown in Table 1. These nights illustrate some of the main features found on a thorough examination of the chosen records. During the display of 8/30/46 the intensity of H β faded between 11:00 and 12:00

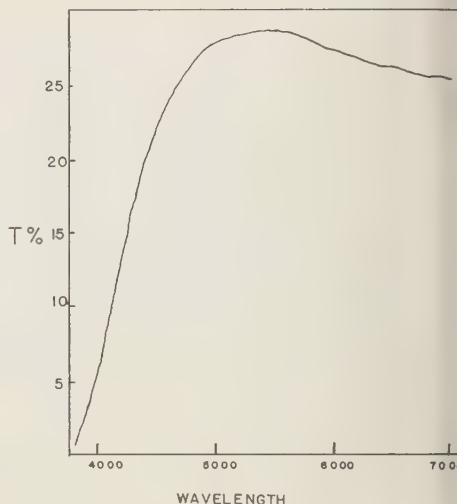


Fig. 2. Spectrograph transmission as a function of wavelength.

P.M., whereas the N₂⁺ bands were reaching their maximum intensity then. During the aurora of 9/17/46, H β remained constant while the N₂⁺ bands were doubling in intensity. The data indicate that there is no constant relationship between any of the N₂⁺ bands considered here and the H β line. (The H β line was chosen in place of the H α line, despite its lower intensity because it is in a clear spectral region, enabling more accurate results to be obtained.)

In the work that follows, the combined intensity of the N₂⁺ bands will be used as a measure of auroral brightness.

Discussion. A detailed examination of the plates shows that the maximum H β intensity occurs before or, at the latest, simultaneously with the maximum intensity of the N₂⁺ bands never afterward. By the time the N₂⁺ bands begin to fade, hydrogen has usually disappeared altogether. If the hydrogen is the initiating agent, the lag in the N₂⁺ emissions indicates metastable levels or slow recombination of nitrogen molecules dissociated by high-energy protons.

One feature of these results was disturbing. The three N₂⁺ bands measured result from transitions from a common excited level of the molecular ion. Therefore, their intensity ratios should be constants [Bates, 1949]. Figure 3 shows

TABLE 1

Date	EST	Visual Record	Intensities, arbitrary units			
			H β	3914	4278	4709
27-28/45	7:00 P.M.	Glow				
	7:41	Faint ray on lower arc				
	7:52	Faint ray on upper arc				
	8:20		1	24	20	3.2
	9:00	Diffuse aurora	2.2	130	50	10
	9:44	Flames				
	9:53	Pulsating surface				
	10:00		0.40	70	22	5.1
	11:00					
	11:25		...	37	9	3.4
	11:45	Rays, diffuse surfaces				
/30-31/46	8:30 P.M.	Glow				
	8:45	Homogeneous arcs, base 45° high				
	9:14-	Rays				
	9:45					
	10:00		0.20	4.2	20	3.3
	10:50	Faint HA, glow to 30°				
	11:00		2.4	40	28	4.3
	11:36	Outburst—rays, diffuse surface				
	11:50	Flames begin				
	12:00		2.0	800	450	22
	1:30 A.M.		1.3	660	420	24
/9-10/48	11:30 P.M.	Glow				
	12:15	Homogeneous arc, base at 10°				
	12:20		0.50	33	19	3.5
	1:00 A.M.	Homogeneous arc at 20°	...	140	61	7.3
	1:07	Rays				
	1:12	Flames				
	1:30	Fades into rayed patches				
	2:00		...	31	18	4.0
	2:45	Great outburst	0.90	319	110	9.0

garithmic plots of these ratios against the total auroral intensity as measured by the sum of the three N $_2^+$ bands considered above. One feature of the plots is believed to be of significance (despite the log-log plots and the large error of each measurement); namely, the distribution of points as the intensity increases is certainly not random. This may be explained on the simple assumption that the auroral forms responsible for the high values were higher in the sky (farther south) than had previously been assumed, resulting in a different correction factor for atmospheric scattering (see appendix). This suggests a correlation between the brightness of an aurora and its southern extent; that is, bright auroras seem to extend farther south than faint ones. (The increase in brightness is too large to be a consequence of the decreased distance from the spectrograph even for a point

source.) That such a relation should exist is not surprising, for *Gartlein* [1944, 1952] has shown the existence of a strong correlation between the southern extent of an aurora and the 3-hour magnetic index, K , and also a correlation between auroral brightness and K number. Thus a southern extent-brightness relation should exist.

If this explanation is correct, it is possible to get an estimate of the relation between auroral brightness and southern extent. (See appendix and Table 2.)

An important feature of this interpretation is that the absolute intensity of an aurora seems to determine its position. The peak of a dim aurora may be at the same latitude as the minimum of a bright one. A given aurora should move south as it brightens.

Miscellaneous observations. During the course

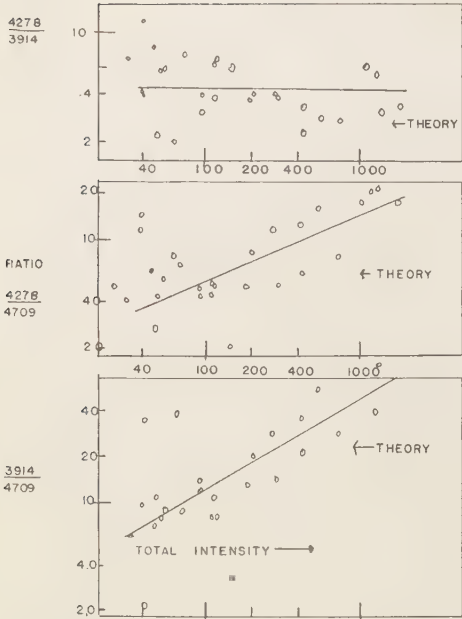


Fig. 3. Nitrogen band intensity ratios as a function of total intensity in all three bands.

of an entire display it was noted that the red [O I] lines reached their maximum intensity before the green [O I] line, but faded more slowly. Also the changes in intensity of the green line during an aurora are relatively greater than the intensity changes of the red lines.

H_{β} was found to be quite strong on some occasions when the visual aurora records indicated that the spectrograph was seeing predominantly active forms. For example, on the night of September 18, 1941, only rayed forms were visible in the field of the spectrograph, yet H_{β} was certainly present [Fan, 1958].

Except as noted above, no relations could be found between the aurora spectrum and magnetic K number, or the spectrum and visual auroral features. The data are probably not sufficiently precise for such relations to appear, if they exist.

For additional information on sequential auroral spectra, from other locations, a list of references is attached.

Conclusions. Although the data are difficult to assess exactly, it is nevertheless possible to draw certain fairly definite conclusions from them.

TABLE 2

$R(I)$	ϕ_A	I/I_{67}	I/I_{67}^*
8	67	1	1
12	65	1.8	1.3
20	63	4.4	2.1†
40	60	14	3.3
65	58	30	3.5

* Corrected for inverse square law assuming point source.

† Aurora touches Ithaca horizon.

1. Proton excitation is not the immediate cause of all auroral emissions. (It may still be the primary source of energy.)
2. Auroral intensity increases with southern extent.
3. The hydrogen emissions are very feeble compared with the total auroral light.
4. The red and green [O I] lines do not develop in phase during an aurora.

APPENDIX

The data presented in the text and Figure 3 show a variation in the ratio of nitrogen band intensities which seems to depend upon auroral intensity. They were obtained by assuming that the auroras observed were on the axis of the spectrograph. In this appendix it will be assumed that the band ratio is a constant, as is strongly indicated by theory, and that the apparent change is due to differential absorption of the atmosphere and a new auroral position. The new position can be determined, and the result is that bright auroras seem to be farther south. Let

- R_A = intensity ratio at aurora, I_1/I_2 , radiation 1 and 2.
 R_{obs} = intensity ratio at Ithaca.
 β_1, β_2 = atmospheric extinction coefficient for radiation 1 and 2, respectively.
 ρ = earth radius.
 d_A, d_E = distance to aurora (A) or estimated (E) position.

Because of the different absorptions of the two

$$R_A = \frac{I_1}{I_2}$$

$$R_{\text{obs}} = \frac{I_1 \exp(-\beta_1 d_A)}{I_2 \exp(-\beta_2 d_A)} \\ = R_A \exp[-d_A(\beta_1 - \beta_2)] \quad (1)$$

$$R_A = R_{\text{obs}} \exp[d_A(\beta_1 - \beta_2)] \quad (2)$$

this can be written

$$= \{R_{\text{obs}} \exp[d_E(\beta_1 - \beta_2)]\} \\ \cdot \exp[(d_A - d_E)(\beta_1 - \beta_2)] \quad (3)$$

the value in the braces is the quantity plotted in Figure 3, on the basis of the assumed position of the aurora on the axis of the spectrograph. (Other procedure was possible.)

Calling these plotted values $R(I)$, and making approximation from the Figure 4 that

$$d_A = \rho(\varphi_A - \varphi_{\text{ith}}) \quad d_E = \rho(\varphi_E - \varphi_{\text{ith}})$$

Equation 3 becomes, solved for φ_A ,

$$\varphi_A = \varphi_E - [12.5/(\beta_1 - \beta_2)\rho] \ln R(I)/R_A \quad (4)$$

The factor 12.5 comes from the fact that the proportion coefficients are for an 8-km atmosphere, whereas the distances refer to an aurora 100 km high: $100/8 = 12.5$; $\varphi_E = 1.08$ radians, assuming the aurora 100 km high and 8.5° up, from Ithaca, New York, and R_A , for 4/4709, theoretically, is 24. Thus

$$\varphi_A = 1.34 - 0.185 \log_{10} R(I) \quad (5)$$

Using this expression gives the approximate relation between geomagnetic latitude and intensity shown in Table 2, columns 1 to 3.

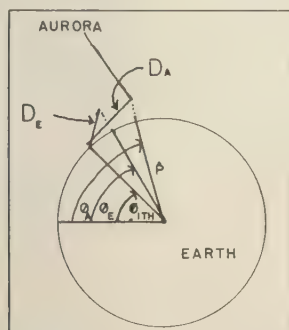


Fig. 4. Geometry of Ithaca auroras.

This estimate has neglected any consideration of the secondary scattering of light into the spectrograph. This introduces some error, since the aurora, as seen from Ithaca, is certainly no point or even line source. On the other hand, the aurora rarely fills more than a quarter of the sky, and fine ray structure can be observed. The primary reason for making no scattering correction is the fact that the aurora moves and changes during the exposure, and no accurate correction is possible. The use of ratios minimizes the error, and the experimental fact is that the ratios change systematically.

Column 4 of Table 2 corrects the results of column 3 for an assumed inverse square law decrease of radiation. Column 3 would apply to an infinite source perpendicular to the line of sight; column 4 would apply to a point source. The actual answer must lie somewhere between these two.

If the aurora is at a higher latitude than 63°N (geomagnetic), part of it is hidden from the spectrograph by the horizon, so the first two entries in Table 2 show an increase due to the fact that the aurora is rising. The last three entries indicate that the intensity is, crudely, proportionate to the decrease in latitude. Although the numbers are only tentative, they are a beginning.

REFERENCES

- Bates, D., The intensity distribution in the Nitrogen band systems emitted from the earth's upper atmosphere, *Proc. Roy. Soc. London, A*, 196, 217-250, 1949.
- Bless, R. C., thesis, Cornell University, 1954.
- Fan, C. Y., Auroral hydrogen emission, *Astrophys. J.*, 12, 420, 1958.
- Gartlein, C. W., Relation of three-hour range index K to auroras seen at Ithaca, New York, *Trans. Am. Geophys. Union*, 25, 533-547, 1944.
- Gartlein, C. W., Appearance of Hydrogen in auroral spectra, *Mem. soc. roy. sci. Liège* 12, 195-198, 1952.
- Van de Hulst, H. C., *The Atmosphere of the Earth and Planets*, edited by G. P. Kuiper, p. 49, University of Chicago Press, 1949.

RELATED SPECTRAL WORK

- Carleton, N. P., and Lawrence, T. R., Absolute cross sections for excitation of nitrogen by protons of a few kev energy, *Phys. Rev.*, 109, 1159-1165, 1958.
- Galperin, G. I., Hydrogen emission and two types

- of auroral spectra, *Planetary and Space Science*, 1, 57, 1958.
- Hunten, D. M., Auroral and night sky spectra, *Ann. géophys.*, 14, 167-174, 1958.
- Montalbetti, R., and W. V. Jones, Hydrogen emissions during auroras over west-central Canada, *J. Atmospheric and Terrest. Phys.*, 2, 43-50, 1957.
- Omholt, A., Studies on the excitation of the aurora borealis, *Geophysica Norvegica*, 20, 1, 1959.
- Romick, G. J., and C. T. Elvey, Variations in intensity of the hydrogen emission during auroral activity, *J. Atmospheric and Terrest. Phys.*, 12, 283-287, 1958.
- (Manuscript received October 14, 1959; revised November 21, 1959.)

A Four-Year Summary of Whistler Activity at Washington, D. C.¹

HAROLD E. DINGER

*U. S. Naval Research Laboratory
Washington 25, D. C.*

Abstract. Whistler and dawn chorus activity as recorded at Washington, D. C., during the period of July 1, 1955, to June 30, 1959, has been tabulated and summarized. Of the 1461 days considered in this analysis, 94 per cent had activity of some form. The equivalent of approximately five hundred 1800-foot reels of magnetic recording tape was analyzed. The last 2 years of the period covered was part of the IGY/IGC Ionospheric Physics Program.

The data presented are the result of a preliminary analysis of the daily log sheets of the Washington Whistler Station for the period July 1, 1955, to June 30, 1959. Some data are available for the years 1953 and 1954 but have not been included because of a difference in programming which would tend to confuse a statistical analysis. A summary of the earlier results appears in a published report [Dinger, 1956]. The IGY portion of the work was done as part of the Whistlers East project, and for purposes of this report is considered as the period from July 1, 1957, to June 30, 1959.

A photograph of the station installed at the Naval Research Laboratory is shown in Figure 1. The equipment was rather conventional, but since loop reception at these frequencies was not feasible in the Washington area because of excessive hum conditions, a long-wire antenna was used for the entire period.

During the first 2 years under discussion the recording equipment was operated automatically for 2 minutes once every 3 hours. During the last 2 years (that is, for the IGY project) recordings were made once each hour. For part of the analysis, however, only every third schedule of the IGY data was included in order to insure uniformity in treating the full 4-year period. For some additional purposes the complete IGY schedule was employed.

In the analysis of records by aural means it was found that the subjective factor could be considerable, particularly as to the assignment of relative whistler strengths and to the

recognition of echoes, multiples, and other non-routine sounds. Fortunately it was possible to arrange for one individual to read almost all the IGY tapes from both the Washington and Bermuda stations. This required over 1168 hours of listening for recorded whistlers, frequently in the presence of high noise levels. Table 1 gives some data on the amount of magnetic tape used for routine recording during the period under discussion. The complete whistler station activities at Washington and Bermuda were maintained with the equivalent of about one and one-third full-time employees. The results given here are, with the exception of a few remarks, confined to those of the Washington station.

During the 4-year period recordings were made while 79 different electric storms were known to have been in progress within a radius of 15 to 20 miles of the station. During 65 (82 per cent) of these periods no whistlers were heard. During 13 (16 per cent) of these periods weak whistlers were heard, and during only 1 period were medium to strong whistlers recorded. Extending this analysis to include storms out to a distance of approximately 100 miles, the figures given in Table 2 were obtained.

There were additional storms in the area, but their times of occurrence were not recorded. Although several investigators [Helliwell and others, 1958; Morgan, 1958] have observed whistlers associated with local lightning flashes, the occurrence has been rare. At Washington, attempts extending over 6 years to connect a whistler definitely to a local lightning flash have

¹Presented at the URSI Fall Meeting, San Diego, October 1959.

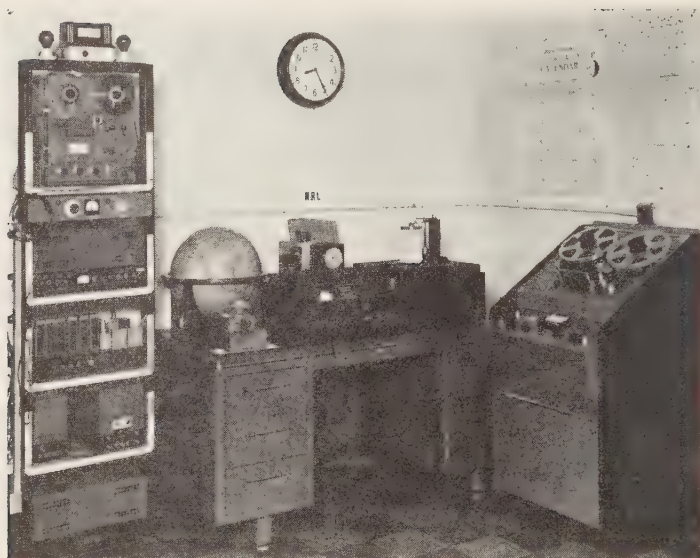


Fig. 1. IGY whistler station at Naval Research Laboratory.

been unsuccessful. Since attempts to associate whistlers received at Washington with storms to the south have also been unsuccessful, possibly it is mostly a matter of geography. A good sferics DF network would be extremely valuable in resolving this question. The question naturally arises whether the high atmospheric noise level is a factor in hearing whistlers dur-

ing local thunderstorm activity. However, weak whistlers and hiss or chorus have been easily detectable between heavy sferics on numerous occasions.

As an example of monthly activity the Washington records for October 1958 contained a count of 901 audible whistlers (with ten schedules missing, which gives just over 30.5 days, or roughly 1 average month). If we multiply 901 by 30 to give an approximate figure for 60 minutes per hour instead of 2 minutes per hour, we get 27,030 whistlers for the complete month. This averages just under 1000 per day. On 2 days during the month no whistlers were recorded. In 1 day the number approached 2800. Table 3 summarizes the recorded activity during the IGY part of the work.² Of the 1461 days considered in this analysis, 94 percent had activity and some form. During the 2-year period there were only 10 days without some form of activity. These percentages would undoubtedly be higher if longer listening pe-

TABLE 1

	1955-1959	IGY
Days	1,461	730
Schedules	23,368	17,520
Reels of tape	487	365
Feet of track	1,753,200	1,314,000
Miles of track	364.6	280

TABLE 2

Number of storms noted	113
Number with medium to strong whistlers	5
Number with weak whistlers	15
Number without local whistler activity	93
Number with whistlers preceding	29
Number with whistlers following	42
Number with strong whistlers following	26

² The following symbols have been used in this report:

W = whistlers. C = chorus.
 E = echoes. R = risers.
 H = hiss. X = other unusual emissions
 M = multiples.

TABLE 3

From the hourly schedules, July 1, 1957, to June 2, 1959:

Whistlers occurred each day for a 5-month period.
Whistlers or VLF emissions occurred each day for a 10-month period.

VLF emissions occurred every hour for a 2-day period.

Out of 730 days:

10 days without any activity.

29 days without any whistlers.

138 days without any H or C.

Of:

98.6% of days with activity.

96.0% of days with whistlers.

81.0% of days with H or C.

oods were used at a site less subject to man-made interference. Stations farther north probably have higher percentages. A 2-day period about the maximum length of time observed at Washington without recorded whistlers.

At some time during every day of one 10-month period extending from September 1957 to June 1958, inclusive, there was activity of some kind—either whistlers, chorus, or both. In only 6 days out of the 303 in this period were no whistlers heard. Including July and

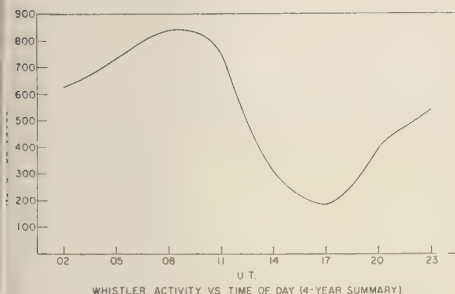


Fig. 2. Diurnal variation in whistler activity.

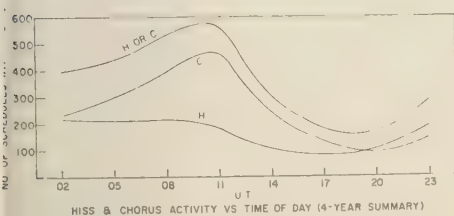


Fig. 3. Diurnal variation of hiss and chorus.

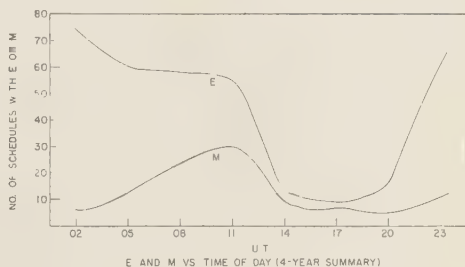


Fig. 4. Diurnal variation of echoes and multiples.

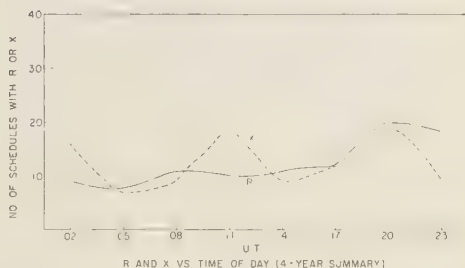


Fig. 5. Diurnal variation of risers and other unusual emissions.

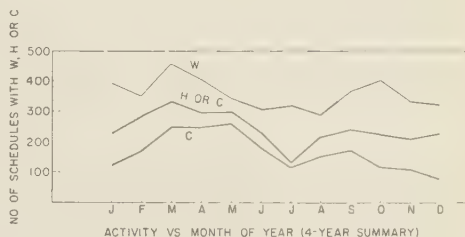


Fig. 6. Month-to-month variation in whistlers, hiss, and chorus.

August of 1957 to complete a full year, only 4 days were without some form of activity, and on 12 days no whistlers were recorded.

Figure 2 is a plot of whistler activity versus time of day totaled over a 4-year period. The shape of the curve is probably fairly representative of the diurnal variation in whistler activity at most of the stations in the network. A plot of the diurnal variation in hiss and chorus activity is shown in Figure 3. Note the much higher peak for chorus than for hiss. The curve for hiss and chorus probably varies much more from station to station than the whistler

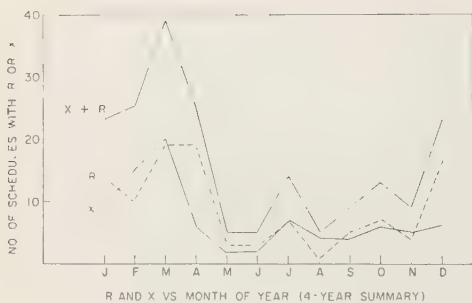


Fig. 7. Month-to-month variation in risers and other unusual emissions.

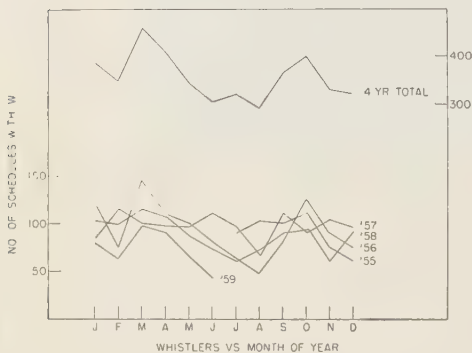


Fig. 8. Month-to-month variation in whistler activity.

curve. Figure 4 gives the diurnal variation in echoes and multiples. The echoes roughly follow the whistler curve. Multiples show a minor peak, but since the multiples have not been separated into the two types, this may not be significant. A plot of the diurnal variation of risers and other unusual emissions is given in Figure 5. Risers peaked late in the afternoon. Other unusual emissions tend to show three peaks. This curve may be somewhat misleading, however, as most of the really interesting displays have rather consistently occurred at about 4 in the afternoon local time.

A 4-year summary of the variation by months of the whistler, hiss, and chorus activity is given in Figure 6. The June-July-August minimum may be due in part to the higher atmospheric noise level and to the longer period of daylight (longer period of higher absorption). This is not the complete explanation, as a care-

ful comparison of records during periods of similar noise levels in July and in March have indicated lower activity in July, although lightning sources are more numerous. A more complete analysis would result by classifying all whistlers as either long and short types. However, relatively few strong short whistlers are recorded in Washington. Chorus and hiss show a greater relative decrease in July than whistlers. A month-to-month plot of riser and unusual activity is given in Figure 7. There appear to be three peaks with the major one consistently about March. Note that a minor peak occurred in July.

A month-to-month plot of whistler activity broken down by years is given in Figure 8.

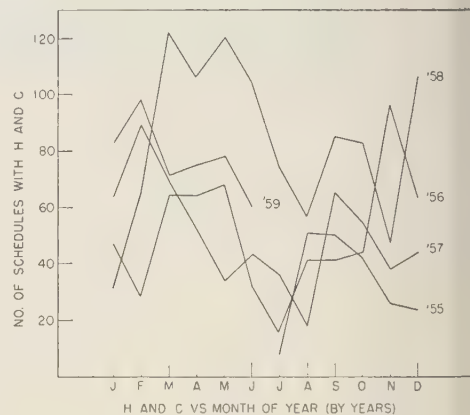


Fig. 9. Month-to-month variation in hiss and chorus activity.

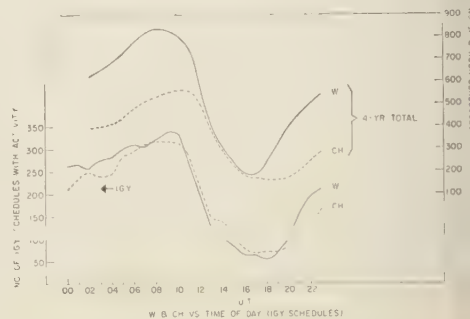


Fig. 10. Diurnal variation of whistlers and combined hiss-chorus activity using the IGY schedules.

ote that whistler activity in 1958 was less than in 1957 except for 3 months. The 1958 average was not highest for any month. Figure 9 shows hiss and chorus for the same period. Most of 1958 was considerably above the corresponding periods in other years.

Figure 10 gives the variation in whistler and combined hiss and chorus activity versus the 24 hourly schedules. The general shape of the curve is not significantly different from that resulting from the use of eight schedules per day.

The limited analysis of the Bermuda station records permits only a few general remarks at this time. As expected, Bermuda whistler and hiss-chorus activity was normally below that at Washington, although at times there was good activity at Bermuda and little or none in Washington. A large percentage of the whistlers recorded at Bermuda were also recorded at Washington. Most, but not all, loud whistlers recorded at Washington were also recorded at

Bermuda. During many periods every whistler recorded at Bermuda was also recorded at Washington. The reverse was seldom true, probably because of the difference in magnetic latitude.

The time of general hiss-chorus activity, as expected, correlated well at Washington and Bermuda, but the discrete periods of activity often differed. At times there was hiss at one place and chorus at the other, although normally it was similar at the two places.

REFERENCES

- Dinger, H. E., Whistling atmospherics, *Naval Research Lab. Rept. 4825*, Sept. 14, 1956.
Helliwell, R. A., W. L. Taylor, and A. G. Jean, Some properties of lightning impulses which produce whistlers, *Proc. IRE*, 46, 1760-1762, 1958.
Morgan, M. G., Correlation of whistlers and lightning flashes by direct aural and visual observation, *Nature*, 182, 332-333, Aug. 2, 1958.

(Manuscript received November 30, 1959.)

The Ion Distribution above the F_2 Maximum

FRANCIS S. JOHNSON

*Lockheed Aircraft Corporation
Missiles and Space Division
Sunnyvale, California*

Abstract. Observations of the ion distribution in the ionosphere indicate that the distribution up to 550 km is controlled in part by changes in the recombination rate with altitude, although diffusion exerts an increasingly important influence above the F_2 maximum. It is further concluded that the ion distribution above 550 km is controlled by diffusion and not by recombination, although a further reduction in recombination rate with increasing altitude does occur. Charge-exchange reactions between oxygen ions and hydrogen atoms at an altitude near the base of the exosphere provide a source of thermal protons which move upward along the magnetic field lines together with an equal number of electrons and produce a medium for the propagation of radio whistlers. The computed distribution of protons and electrons is in reasonable agreement with whistler observations. The whistler medium is of telluric origin, and the source of ionization is the same as that for the F region. It is suggested that the region above 1800 km in which protons predominate over all other ions be called the protonosphere, to distinguish it from the lower ionized region normally referred to as the ionosphere.

Introduction. The physics underlying the formation of the D , E , and F_1 regions of the ionosphere has been discussed by Chapman [1931] in his classic theory of ion layer formation. Although simplifying assumptions were made, such as constant scale height and constant recombination coefficient, the basic features of the lower ionospheric layers are generally well described by his theory. A characteristic of the Chapman layer is that the ionizing radiation causing it is most strongly absorbed at the peak of the layer. The F_2 region is not a Chapman layer, and the physics underlying this layer was largely brought out by Bradbury [1938], who suggested that decreasing electron loss coefficient with altitude caused the F region to split into two layers, the F_1 at the level of maximum absorption of the least penetrating of the solar ionizing radiation, and the F_2 at some higher level controlled by the decreasing loss coefficient. Bradbury envisioned the variation in electron loss coefficient with height as due to variation in the rate of negative-ion formation; later, Havens, Friedman, and Hulburt [1955] showed that the mechanism which produced the variation in electron loss coefficient with height was the decreasing molecular-oxygen concentration with altitude, with

a charge-exchange reaction controlling the actual loss rate. As this explanation apparently is not doubted, it is appropriate to refer to the F_2 region as a Bradbury layer, since the mechanism for the formation of the layer is generally in agreement with Bradbury's hypothesis.

Ferraro [1945] has discussed diffusion in the ionosphere, especially in connection with the F_2 region. In his investigation, he concluded that the neutral-particle density at the F_2 peak must exceed a stated value; otherwise, diffusion would cause the F_2 peak to disappear by downward diffusion faster than it forms. Although it was not explicitly so stated, Ferraro's investigation implies that the distribution of ions above the F_2 maximum is governed by diffusion. In this paper, the role of diffusion in governing the distribution of ions above the F_2 maximum will be considered.

The propagation of radio whistlers indicates the existence of an ionized medium far above the normal ionosphere, extending upward a distance of several earth radii [Storey, 1953]. Although the source of this medium has not been critically examined, it has most often been thought of as of solar origin. It will be argued here that the medium consists of protons and

electrons which originate near the base of the exosphere by charge exchange and which above this altitude are in hydrostatic equilibrium. The name protonosphere is adopted here to describe the ionized medium above about 1800 km where protons are the principal ionized constituent, the name ionosphere being reserved for the lower region consisting of heavier atmospheric ions, such as atomic oxygen.

Diffusion. Somewhere above the F_2 maximum, diffusion becomes so rapid that it, rather than recombination, becomes the dominant factor controlling the ion distribution. Above the altitude where diffusion becomes dominant, the ions are distributed according to a diffusive equilibrium; i.e., each ionic species is distributed according to its own scale height, independently of the neutral particles and of the other ionic species. Consider first the case of a single ionic species. If the electron temperature is T_e , and the ion temperature is T_i , the scale height of the combination of ions and electrons, which is constrained to remain very nearly electrically neutral, is given by

$$H_i = k(T_i + T_e)/(m_i + m_e)g \quad (1)$$

$$\approx k(T_i + T_e)/m_i g$$

where k is Boltzmann's constant, m_i is the mass of an individual ion, m_e is the electron mass, and g is the acceleration of gravity. When several ionic species are present, equation 1 applies separately for each.

The ions, upon formation by charge exchange, and especially the electrons, upon formation by photoionization, may have considerably more than thermal energy. The ions quickly adjust their temperature to that of the neutral particles, since they can lose energy effectively in colliding with neutral particles. The electrons with energies less than a few ev do not lose energy very quickly in colliding with atoms or ions, because of the large mass difference. However, the newly released and energetic electrons quickly share their energies with other electrons. The possibility therefore exists that the electron temperature may be greater than the ion temperature and the neutral-particle temperature, while the latter two are certainly very nearly equal. When, as occurs in some types of plasma experiments, the electron tem-

perature is substantially greater than the ion and neutral-particle temperature, the electrons are described as having a runaway temperature. As the total amount of energy supplied to the electrons by photoionization is increased, the electron temperature rises somewhat above the ion and neutral-particle temperature; when a critical value of energy input is reached, the electron temperature begins to rise rapidly, and with the higher average thermal velocity, the electrons become even less effective in transferring their excess energy to ions because of decrease in collision cross section. This is the condition leading to the runaway electron temperature. In the F_2 region, the absorption of solar energy is close to this critical value; on the basis of several sorts of evidence, however, *Hanson and Johnson* [1960] conclude that the rate of absorption of solar energy is below the critical value. Hence, we will assume here that the ion, electron, and neutral-particle temperatures are all equal.

The expression for the scale height of the ionosphere assumes a simpler form if it is assumed that $T_i = T_e = T$ and that the ions and neutral particles have the same mass. The scale height for the ions and electrons is then

$$H_i = 2kT/mg = 2H \quad (2)$$

where H is the scale height of the neutral particles. This equation indicates that the scale height for the ions is just twice that for the neutral particles. Thus, when making hydrostatic calculations, it is convenient to regard the ion-electron gas as if all the particles had the same mass—just half the actual ion mass. It must be kept in mind, however, that a runaway electron temperature would require that equation 1 be used instead of equation 2, and that the ion mass would have to be reduced by a factor larger than 2 when making hydrostatic calculations.

The equation describing the time rate of increase of ion concentration at any level, including the effect of diffusion, has been given by *Ferraro* [1945] as

$$\frac{\partial n_i}{\partial t} = q - \alpha n_i^2 + \frac{\partial}{\partial h} \left\{ D \frac{\partial n_i}{\partial h} + D \frac{n_i}{2H_i} \right\} \quad (3)$$

where n_i is the ion concentration, q is the rate of formation of ions, α is the recombination

efficient, h is the height variable, and D is the diffusion coefficient. Martyn [1959] has given a similar equation which includes in addition the effects of vertical motion and of velocity divergence of the ions, which we disregard here. The expression in braces in equation (3) gives the downward flow of ions at any level due to diffusion. The effect of the magnetic field is not included explicitly, but should be taken into account whenever the ion mean free path is large compared with the ion cyclotron radius by reducing the diffusion coefficient by the factor $\sin^2 I$, where I is the dip angle of the magnetic field. Although Ferraro gave values for the diffusion coefficient D , more recent and smaller values have been proposed by Dalgarno [1957].

Recombination or loss rate. It is generally accepted that the rate of loss of ions in the F_2 region is more nearly proportional to the first power than to the square of the ion concentration. This is in agreement with the idea that O^+ ions are formed by photoionization, and that they undergo charge exchange with molecular oxygen to form molecular oxygen ions which in turn dissociatively recombine with electrons [Havens, Friedman, and Hulburt, 1955]. The charge-exchange reaction is the one that determines the rate of electron loss. The loss-rate coefficient decreases with altitude since it is controlled by the concentration of molecular oxygen. Even though it is accepted that the loss-rate coefficient depends linearly upon the ion or electron concentration and upon the molecular oxygen concentration, it is often convenient to consider an effective recombination coefficient which has the value that would be required to explain the rate of electron loss if the loss were properly described by a recombination coefficient.

The electron loss-rate coefficient is important in establishing the level above which diffusion is dominant in controlling the ion distribution, since the loss-rate coefficient determines the rate of a competing process. However, if the rate of electron loss can be described by a recombination coefficient or an effective recombination coefficient that does not change with altitude in a region where the absorption of the ionizing radiation is weak, then diffusion will not affect the distribution, for the equilib-

rium ion densities are the same as those that would result from a diffusive equilibrium. This is the condition that would prevail in the upper part of a Chapman layer, well above the peak ionization. The rate of production of ions at the level h is

$$q(h) = I(h)n\beta \\ = I_0 n\beta \exp\left(-\beta \int_h^\infty n \, dh\right) \quad (4)$$

where n is the number density of atmospheric particles at the level h , β is the absorption cross section, and I_0 is the intensity outside the atmosphere of the ionizing radiation that produces the F_1 region. At the F_2 peak and above, $I(h)$, the intensity of ionizing radiation, is nearly constant with altitude h and is nearly equal to I_0 , since the exponential term in equation 4 is nearly unity there. Therefore, equation 4 becomes

$$q(h) \approx I_0 n\beta$$

For photoionization equilibrium,

$$q - \alpha n_i^2 = 0$$

where α is the recombination coefficient and n_i is the concentration of ions. Hence,

$$n_i = \{[I(h)\beta/\alpha]n\}^{1/2} \\ = \{[I(h)\beta/\alpha]n_0\}^{1/2} e^{-h/2H} \quad (5)$$

where n_0 is the concentration of neutral particles at the level from which h is measured. This equation shows that the ions are distributed according to a scale height just twice that of the neutral particles, which is the same distribution that would result from diffusive equilibrium. This result is implicitly included in the formula for a Chapman layer; well above the ion peak, the scale height for the ions in a Chapman layer is twice the neutral-particle scale height.

We may now specify two conditions that must be fulfilled in order for a distribution of ions to differ from a diffusive equilibrium distribution: (1) the recombination coefficient must exhibit a significant variation with altitude in the region over which the distribution is to differ from a diffusive equilibrium; (2) diffusion must proceed slowly enough so that it will

not establish a diffusive distribution in spite of the variations in recombination coefficient. Both the above conditions must be met in the vicinity of the F_2 maximum. The first condition is seen to be satisfied; a recombination coefficient between 10^{-10} and 10^{-11} $\text{cm}^3 \text{sec}^{-1}$ is generally accepted as required at the F_2 maximum, whereas a smaller value, 10^{-12} $\text{cm}^3 \text{sec}^{-1}$, presumably applies at some higher level [Havens, Friedman, and Hulbult, 1955]. The second condition was discussed by Ferraro, and as was mentioned in the previous section it sets a lower limit on the atmospheric density at the altitude of the F_2 maximum.

The ion distribution. Measurements of the ion distribution through the F_2 peak and up to an altitude of 550 km were made by Farmer and Robinson [1959] during an Atlas rocket flight from Cape Canaveral. These measurements indicate that a diffusive equilibrium distribution had not become established up to that altitude. Since the measurements at different altitudes were made at different horizontal ranges rather than in a vertical sounding, there is a reasonable probability that variations in the horizontal structure of the ionosphere somewhat distorted the observation of the vertical structure. If the electron distribution of Farmer and Robinson is correct in indicating that diffusive equilibrium was not established up to the altitude of 550 km, the diffusion coefficient published by Dalgarno [1957] must be too large by a factor of about 5. However, only a small modification of Farmer and Robinson's distribution would be required to bring it into agreement with diffusive equilibrium above 450 km, in which case Dalgarno's diffusion coefficient would fit fairly well; the required change in distribution is approximately a 10 per cent increase near 450 km. The presence of 2 to 3 times as many electrons above the F_2 maximum as below, deduced from the Faraday rotation of lunar echoes [Evans, 1957], indicates either that the electron concentration immediately above the F_2 maximum falls slowly for 100 or 200 km before diffusion becomes so rapid as to control the distribution or that the electron temperature above the F_2 maximum is a factor of 2 or 3 higher than the neutral-particle temperature. One or the other of these conditions must be fulfilled to account for the

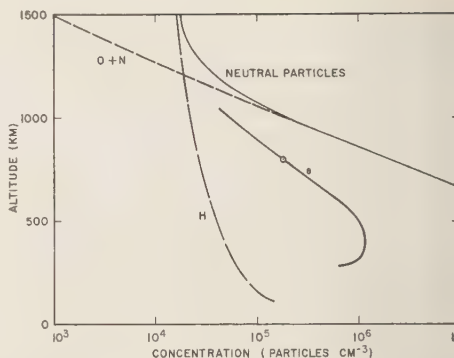


Fig. 1. Electron and neutral-particle concentration in the ionosphere and exosphere.

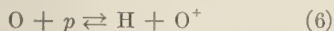
presence of so many electrons above the F_2 maximum, and as the latter condition, high electron temperature, seems inadmissible [Hansen and Johnson, 1960], the lunar echo work seems to favor a distribution about as presented by Farmer and Robinson.

The ion distribution and the neutral-particle distribution through the ionosphere are shown in Figure 1. The ion distribution in the F_2 region and up to 550 km is that obtained by Farmer and Robinson [1959]; the concentration at the F_2 peak has the rather high value 1.15×10^6 ions cm^{-3} . A decrease in recombination coefficient by a factor of about 10 between the F_2 peak and the 550-km level is consistent with the observed ion concentration curve. An additional decrease in the recombination coefficient above the 550-km level should occur, the value finally falling to about 10^{-12} $\text{cm}^3 \text{sec}^{-1}$. Above 550 km, where a diffusive equilibrium must prevail, the scale height for the ions is taken to be twice that of the neutral particles. The plotted point at 795 km is a value measured in Sputnik III [Krassovsky, 1959]. The neutral-particle curve in Figure 1 is based up to 1000 km on the analysis by Harris and Jastrow [1959] of the Vanguard 1958 β_2 orbit decay; this distribution indicates a temperature of 1250°K in the F_2 region and exosphere, assuming a molecular weight of 16. Above 1000 km, the particle densities are modified by the presence of a hydrogen cloud around the earth. The existence of this cloud was first explicitly noted by Chapman [1957]. Quantitative values

it have been put forth by *Singer* [1959a and b] on the basis of the calculated lifetimes of Van Allen particles and by *Johnson* [1960] on the basis of scattered hydrogen Lyman- α radiation observed above the E region at night. Johnson's values, which are about two decades lower than those proposed by Singer, are used here, as they are based on a much more reliable normalization procedure.

The total number of ions per unit column for the distribution shown in Figure 1 is 4.25×10^{18} ions cm^{-2} , with $2\frac{1}{2}$ times as many ions above the F_2 maximum as below. This ratio for ions above the F_2 peak to those below is in good agreement with the Faraday rotation measurements, although a larger factor is sometimes favored. A small addition to the ions above the F_2 maximum is contributed by ionized hydrogen in the protonosphere, as will be seen below.

The Source of protons. Hydrogen is the most important constituent of the atmosphere above 1300 km, and it is the presence of hydrogen in the atmosphere that gives rise to the medium which makes possible the propagation of radio whistlers. Hydrogen ions are formed by charge exchange between hydrogen atoms and oxygen ions [Dungey, 1955]. The ionization potentials are almost identical, the difference being only a small fraction of the thermal energy; hence, the charge-exchange reaction



should proceed equally rapidly in both directions. Since the cross sections for collisions between ions and neutral particles are generally only slightly larger than the cross section for collisions between neutral particles, the importance of this reaction in determining the equilibrium relative concentrations of the two ions in the exosphere should be greatest at a level near the base of the neutral-particle exosphere. The base of the exosphere is the significant level, because collisions between protons and oxygen atoms, and between oxygen ions and hydrogen atoms, will be relatively rare above the base of the exosphere. At higher levels, the distribution of oxygen and hydrogen ions will be in diffusive equilibrium.

The base of the neutral-particle exosphere is at an altitude of 550 km, as determined from

the Vanguard 1958/32 orbital decay data. This level can be defined as the most probable level of penetration for atmospheric particles when they fall vertically downward into the atmosphere. If the descending particles have an isotropic velocity distribution, as they generally do, their most probable altitude of penetration is 0.70 scale height above the base of the exosphere, or an altitude of about 600 km. If the collision cross sections for protons and oxygen atoms, and for oxygen ions and hydrogen atoms, are the same as those for neutral atmospheric particles, 600 km is the most probable level for penetration of protons before colliding with oxygen atoms, and the most probable level of origin for protons found moving up through the lower exosphere after formation in a charge-exchange reaction between oxygen ions and hydrogen atoms. The cross sections for ions colliding with ions is much larger than the cross sections mentioned above, and as a result, there is no ion exosphere, i.e., no level above which collisions among ions can be neglected. Even for a concentration as low as 500 protons cm^{-3} at 1250°K, the mean free path is about 1000 km, or much less than the scale height for protons.

The relative concentration of hydrogen and oxygen atoms at an altitude of 600 km controls the relative concentration of protons and oxygen ions at that level and thus controls the concentration of protons and electrons throughout the exosphere. Since the difference in the ionization potentials is only a small fraction of the thermal energy, the charge-exchange reaction should proceed equally rapidly in either direction, and it is reasonable to expect that the neutral particle and ion ratios will be equal, or that

$$[\text{H}]/[\text{O}] = [p]/[\text{O}^+] \quad (7)$$

At an altitude of 600 km, $[\text{H}] \approx 4 \times 10^4 \text{ cm}^{-3}$ and $[\text{O}^+] \approx 6 \times 10^6 \text{ cm}^{-3}$. The neutral-particle concentration is $1.5 \times 10^7 \text{ cm}^{-3}$; the relative concentration of oxygen and nitrogen atoms depends on the degree of nitrogen dissociation and the region in which the dissociation occurs. A proton concentration at 600 km of about $3 \times 10^8 \text{ cm}^{-3}$ is required to produce the whistler medium, and according to equation 7 this requires an oxygen concentration of 0.8×10^7

cm^{-3} ; since the total number of particles at this altitude is $1.5 \times 10^7 \text{ cm}^{-3}$, the implication is that the atomic nitrogen concentration is about the same as that of atomic oxygen. Hertzberg [1958, 1959] has pointed out that nitrogen dissociation is to be expected above about 200 km, owing to the ion-atom interchange reaction,



and this is roughly consistent with equal concentrations of atomic oxygen and atomic nitrogen near 600 km.

The reaction cross section for the charge-exchange reaction may be large compared with neutral-particle collision cross sections, since it commonly is with resonant reactions; this would result in a level higher than 600 km being dominant in controlling the relative ion concentrations in the exosphere. A higher proton concentration would then result, because the ratio of the concentration of oxygen ions to that of oxygen atoms increases with altitude, whereas the hydrogen atom concentration is nearly constant. On the other hand, if the reaction cross section for charge exchange were somewhat smaller than the collision cross section, the 600-km level would still remain the controlling region for determining the equilibrium concentration of ions. The effect of the small charge-exchange reaction cross section would only be to lengthen the time required for the establishment of the equilibrium relative concentrations without actually changing the equilibrium values. Measurements of the reaction cross section for the reaction $\text{O}^+ + \text{H} \rightarrow p + \text{O}$ have been made with ion energies above 400 ev, and a value greater than that for the gas kinetic collision cross section was obtained (R. F. Stebbings, private communication). This result can not readily be extrapolated to thermal energy, but it indicates that there is no problem in releasing enough protons to produce the protonosphere.

The ion distribution in the protonosphere. In computing the distribution of protons above the base of the exosphere, a hydrostatic equation can be used. One must decide whether the presence of magnetic field affects the distribution. Since the mean free paths for the protons and electrons are long compared with the cy-

clotron radii, the average motion of the protons and electrons is along the magnetic field lines, and the mean motion of the particles is not isotropic. However, collisions among the protons and electrons remain rather frequent when the temperature is as low as 1250°K; even for concentrations as low as 300 particles cm^{-3} , the collision frequency is about $0.3 \times 10^{-9} \text{ sec}^{-1}$. Under these conditions, the detailed velocity distribution remains isotropic, even if the mean motion of individual particles is not. As long as the detailed velocity distribution remains isotropic, the distribution of ions must be in accordance with the hydrostatic relationship [Spitzer, 1952].

The only correction that must be made to the generalized hydrostatic equation is that due to the rotation of the earth, since all the ions and electrons, wherever they may be in the magnetic field, must rotate with the earth about the earth's axis. The modified hydrostatic equation for the distribution of ions or electrons in the exosphere, or in a region where diffusive equilibrium prevails, is

$$N(h) = N_0 \exp [g_0 R_0^2 m / (R_0 + h) k T] \cdot \exp [m \Omega^2 (R + h)^2 \cos^2 \theta / 2 k T] \quad (8)$$

where $N(h)$ is the number density at altitude h , N_0 is a normalization constant, g_0 is the acceleration of gravity at the earth's surface, R_0 is the radius of the earth, m is half the ion mass, Ω is the rotational velocity of the earth about its axis, θ is the geocentric angle between the equatorial plane and the point considered (i.e., latitude), and k is Boltzmann's constant. The second exponential is due to the centrifugal force. At about 5 earth radii near the equatorial plane, the centrifugal term due to the earth's rotation is as great as the gravitational attraction. Since the two effects oppose each other, the distribution of protons with altitude is constant in this region.

The distribution of protons and electrons based upon equation 8 is shown in Figure 2 for the region above 550 km. The distribution applies along magnetic lines of force, and for lines leaving the earth's surface at latitudes lower than 65° the effect of $\cos \theta$ in the second exponential term in equation 8 is small. Therefore, the distribution shown in Figure 2 applies

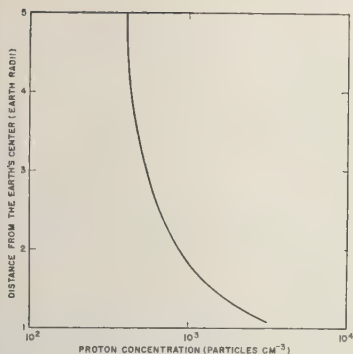


Fig. 2. Proton distribution near the earth.

ell throughout the whistler medium. The model is in good agreement with whistler data. The ion and proton concentrations in the F region and above are shown in Figure 3. The lower part of the proton curve falls to small concentration values because of the effectiveness of atomic oxygen at the lower altitudes in changing protons to hydrogen atoms without a corresponding increase in the supply of protons by reactions between oxygen ions and hydrogen atoms. The protons contribute an additional 0.2×10^{12} ion cm^{-3} to the total ion concentration in a unit column through the ionosphere. The neutral-particle concentration curve is also included in Figure 3.

Because of the existence of the neutral hydrogen cloud around the earth, it is concluded that the protons associated with whistler propagation must be of thermal energy, irrespective of their source. If it is hypothesized that the protons do not originate in the charge-exchange reaction, but that they arrive from some other source and are trapped in the magnetic field with greater than thermal energy, then, with a 24-hour time constant, the protons would be reduced to thermal energy. This reaction would occur because of the charge-exchange reaction between hydrogen atoms and protons, for which *Fite and others* [1958] have measured the cross section. This argument has been applied by *Stuart* [1959] to show that the Van Allen radiation cannot contain protons with energies less than 50 kev, and by *Dessler and Parker* [1959] to explain the decay of the main phase of magnetic storms. Once reduced

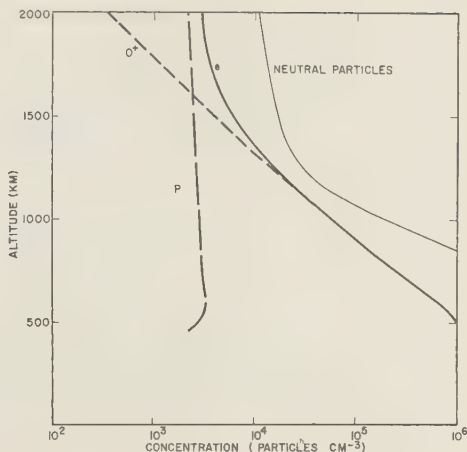


Fig. 3. Ion and neutral-particle concentration in the ionosphere and exosphere.

to thermal energies, the protons would quickly develop an isotropic distribution of velocities and come into equilibrium with the oxygen and hydrogen concentrations at the base of the exosphere, and the distribution would be reduced to the form presented here. Therefore, it is concluded that there is no alternative but to accept the hypothesis that the bulk of the ions associated with whistler propagation originate in the upper part of the F region.

It is of interest to note that the Van Allen radiation represents a vanishingly small percentage of the total ion concentration and its contribution to the protonosphere can be ignored. For example, the inner zone of the Van Allen radiation belt contains about 10^{-6} proton cm^{-3} with energies greater than 40 Mev and about 10^{-1} electron cm^{-3} with energies greater than 20 kev [*Van Allen and Frank*, 1959].

Conclusions. In the upper part of the F region, above the F_2 maximum, recombination of ions is important up to 450 or 550 km in controlling the distribution of ions, with a decrease of perhaps a factor of 10 in the recombination coefficient occurring between the F_2 peak and 550 km. Diffusion is rapid enough above 550 km so that the ions and neutral particles are distributed according to a diffusive equilibrium, although the recombination coefficient probably exhibits further decrease with increasing altitude above 550 km.

The ions above 1800 km are predominantly protons, and those below, atomic oxygen, the transition region being several hundred kilometers thick. The protons originate by charge exchange between neutral hydrogen atoms and oxygen ions; hence, the source of the ionization is the same as that for the F region. The protons are distributed vertically according to a hydrostatic law, in which the gravitational attraction is modified by centrifugal force, especially at the higher levels. The electrons that are present in the same concentrations as the protons provide the medium for the propagation of whistlers. Since the plasma is of telluric origin, whistlers do not yield any information about interplanetary gas or the solar corona near the earth's orbit.

REFERENCES

- Bradbury, N. E., Ionization, negative-ion formation, and recombination in the ionosphere, *Terrestrial Magnetism and Atmospheric Elec.*, **43**, 55-66, 1938.
- Chapman, S., The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, *Proc. Phys. Soc. London*, **43**, 26-45, 1931.
- Chapman, S., Speculations on the atomic hydrogen and the thermal economy of the upper atmosphere, *The Threshold of Space*, edited by M. Zelikoff, Pergamon Press, London, pp. 65-72, 1957.
- Dalgarno, A., Ambipolar diffusion in the F_2 Layer, *J. Atmospheric and Terrest. Phys.*, **12**, 219-220, 1957.
- Dessler, A. J., and E. N. Parker, Hydromagnetic theory of geomagnetic storms, *J. Geophys. Research*, **64**, 2239-2252, 1959.
- Dungey, J. W., Electrodynamics of the outer atmosphere, *The Physics of the Ionosphere, Report of 1959 Cambridge Conference*, The Physical Society, London, p. 406, 1955.
- Evans, J. W., The electron content of the Ionosphere, *J. Atmospheric and Terrest. Phys.*, **11**, 259-271, 1957.
- Farmer, D. J., and W. A. Robinson, Density distribution of free electrons in the upper atmosphere, Paper presented at International Scientific Radio Union meeting, Washington, D. C. May 4-7, 1959.
- Ferraro, V. C. A., Diffusion of ions in the ionosphere, *Terrestrial Magnetism and Atmospheric Elec.*, **50**, 215-222, 1945.
- Fite, W. L., T. R. Brackman, and W. R. Snodgrass, Charge exchange in proton-hydrogen-atom collisions, *Phys. Rev.*, **112**, 1161-1169, 1958.
- Harris, I., and R. Jastrow, Atmosphere derived from rocket and satellite data, *Planetary and Space Science*, **1**, 20-26, 1959.
- Hanson, W. B., and F. S. Johnson, Electron temperatures in the F region, in preparation, 1959.
- Havens, R. J., H. Friedman, and E. O. Hulbut, The ionospheric F_2 region, *The Physics of the Ionosphere—Report of 1954 Cambridge Conference*, The Physical Society, London, pp. 23-24, 1955.
- Hertzberg, M., Atomic nitrogen production in ion-atom interchange reactions in the upper atmosphere, *J. Geophys. Research*, **63**, 856, 1958.
- Hertzberg, M., Ion-atom interchange as an important source of ionospheric atomic nitrogen, *J. Geophys. Research*, **64**, 1106, 1959.
- Johnson, F. S., The telluric hydrogen corona, *Astrophys. J.*, March 1960.
- Krassovsky, I. S., Exploration of the upper atmosphere with the help of the third Soviet sputnik, *Proc. IRE*, **47**, 289, 1959.
- Martyn, D. F., The normal F region of the Ionosphere, *Proc. IRE*, **47**, 147-154, 1959.
- Singer, S. F., Distribution of neutral hydrogen in the earth's exosphere, *Bull. Am. Phys. Soc. II*, **222**, 1959a.
- Singer, S. F., Artificial modification of the earth's radiation belt, *Advances in Astronaut. Sci.*, **1**, 335-354, Plenum Press, New York, 1959b.
- Spitzer, L., Ideal plasma, *Astrophys. J.*, **116**, 219, 1952.
- Storey, L. R. O., An investigation of whistling in the ionosphere, *Phil. Trans. Roy. Soc. London, Ser. A*, **246**, 113-141, 1953.
- Stuart, G. W., Satellite-measured radiation, *Phil. Mag. Lett.*, **2**, 417, 1959.
- Van Allen, J. A., and L. A. Frank, Radiative measurements to 658,300 km with Pioneer 1, *Nature*, **184**, 219, 1959.

(Manuscript received October 30, 1959.)

Abnormal Features of the F_2 Region of the Ionosphere at some Southern High-Latitude Stations

R. G. RASTOGI¹

*Radio and Electrical Engineering Division
National Research Council
Ottawa, Canada*

Abstract. The paper describes some abnormal features of the F_2 layer at high-latitude stations in the southern hemisphere. The variation of the midday value of the critical frequency of the F_2 layer (f_oF_2) with magnetic dip shows asymmetry between the northern and southern stations of the west (American) zone, but not of the east zone. The control by the earth's magnetic field is indicated in the latitudinal distribution of even the midnight values of f_oF_2 . The diurnal variation of f_oF_2 at Port Lockroy shows abnormal minimum at midday and maximum at midnight during the summer months. These abnormal features of the F_2 layer at Port Lockroy are explained on the basis of the horizontal transport of ionization as guided by the earth's magnetic field.

Introduction. Rastogi [1959a] has shown that the latitudinal variation of noon critical frequencies of the F_2 layer for any month, as well as the diurnal variation for any season and the seasonal variation of noon values of f_oF_2 at different stations, could be easily interpreted in terms of their true magnetic and not the realized geomagnetic latitudes. However, it has been noticed that the values of f_oF_2 at southern stations of the west zone were lower than those of the corresponding northern stations, the difference being greatest during local summer months. Some abnormal features of the F_2 region at these stations are described, and possible explanations are suggested.

Latitudinal variation of midday f_oF_2 in the west and east zones. First, the effect of the magnetic dip on the noontime critical frequency of the F_2 layer in the northern and southern hemispheres is compared separately for the west zone and the east zone. Figure 1 shows the relationship between midday f_oF_2 and magnetic dip for the months typical of each season in the low-sunspot year 1954. The curves drawn through the points corresponding to the northern or southern stations of the east zone are roughly alike to indicate that there is no discrepancy between the two hemispheres in the

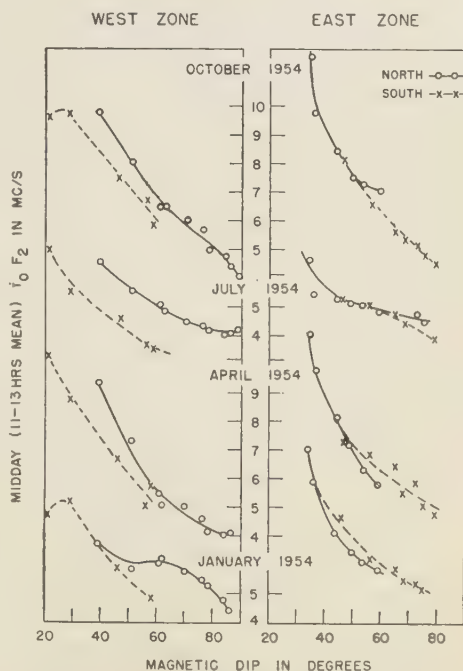


Fig. 1. Noon critical frequency of the F_2 layer during each season as a function of magnetic dip for the west (American) zone and east (Australian) zone. Note the discrepancy between the northern and southern stations in the west zone but the absence of it in the east zone.

¹ National Research Council Post-Doctorate Fellow.

east zone. The lines joining the points corresponding to the northern or southern stations of the west zone differ appreciably from each other; the two curves are well separated. The critical frequency is, in general, lower at the south than at the north station having the same magnetic dip. This discrepancy between the two hemispheres is greatest during July and least in January.

Previous investigations have failed to show any such discrepancy between the northern and southern hemispheres. *Lepechinsky and Pillet*. [1950] pointed out that the distributions of noon f_oF_2 show better symmetry between the northern and southern hemispheres when plotted against magnetic inclination than when plotted against geographic or geomagnetic latitude. *Shapley* [1956] compared the relationship between noon f_oF_2 and magnetic dip at the northern and southern hemispheres, concluding that no clear discrepancies were present between the two hemispheres. The diagrams shown by *Lepechinsky and Pillet* and by *Shapley* had incorporated only one southern west zone station, Falkland, and so the abnormally low ionizations in the F_2 region in that region could not be detected.

It is concluded that the F_2 critical frequencies for the noon hours at southern stations of the west zone are lower than at corresponding northern stations or at southern stations in other zones.

Latitudinal variations of midnight f_oF_2 . Next, the latitudinal variations of the midnight critical frequencies of the F_2 layer are examined for the two solstices.

Figure 2 shows the variation of midnight f_oF_2 against the geographic, geomagnetic, and magnetic latitudes for the period from November 1953 to February 1954. It is interesting to find that the critical frequency continues to increase as one proceeds from the northernmost station to the southernmost station of the west zone, whereas in other zones the maximum of f_oF_2 occurs at middle south latitudes. The values of f_oF_2 at southern stations of the west zone are much higher than at other southern stations having the same geographic or geomagnetic latitude. The greatest discrepancies occur at Falkland, Deception, and Port Lockroy. Comparing the dispersion of points in the three

plots one notices that the points are least scattered in the magnetic latitude plot. If the magnetic latitude is assumed as the coordinate most effective in controlling the distribution of midnight f_oF_2 , there does not seem to be much difference between the west zone and other zones. The critical frequencies at Falkland and Deception are comparable to the maximum value of f_oF_2 at other zones. Any decrease of f_oF_2 south of 30° magnetic latitude in the South American zone cannot be ascertained, owing to the absence of any ionospheric station south of Port Lockroy during the period under study.

Figure 3 shows the latitudinal variations of midnight f_oF_2 averaged for the solstitial months May to August 1954. There does not seem

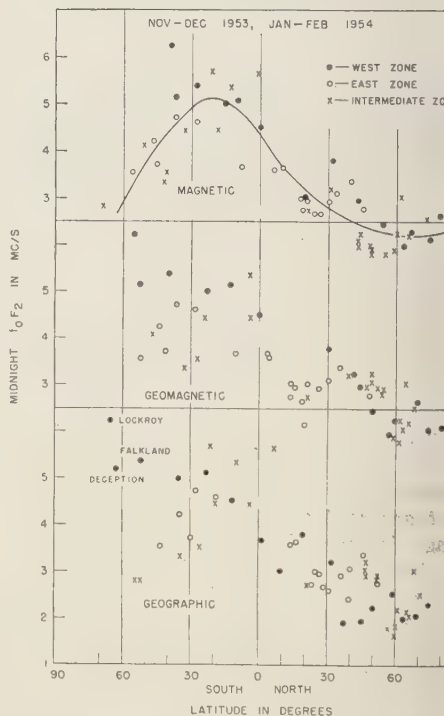


Fig. 2. Midnight critical frequency of the F_2 layer during the southern solstitial months (November 1953 to February 1954) as a function of geographic, geomagnetic, and magnetic latitude. Note that the scatter of points is least in the magnetic latitude plot, indicating a significant magnetic control on nighttime distribution of f_oF_2 .

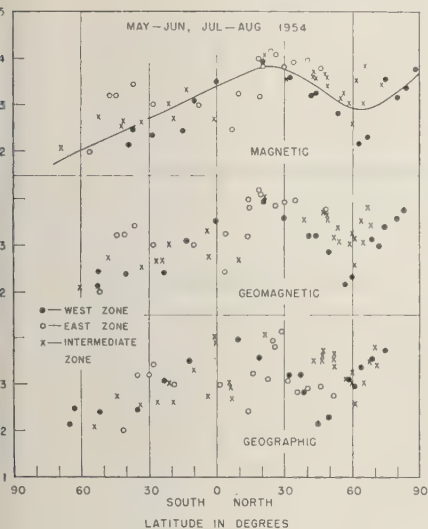


Fig. 3. Midnight critical frequency of the F_2 layer during the northern solstitial months (May-August 1954) as a function of geographic, geomagnetic, and magnetic latitude. Note that the scatter of points is greatest in the geographic latitude plot, indicating the geomagnetic control over the distribution of f_oF_2 even during the night.

any significant discrepancy in the variations of f_oF_2 in the different zones.

It is thus concluded that the earth's magnetic field (true magnetic latitude based on ground measurements) has a significant control over the distribution of midnight critical frequencies of the F_2 layer.

Diurnal characteristics of F_2 layer at South American stations. The diurnal variations of f_oF_2 at Falkland, Deception, and Port Lockroy during the winter month of July of years with low (1954) and high (1956) sunspot number are compared in Figure 4. The critical frequency at each of these stations for July 1954 shows a maximum at midday. The variation of f_oF_2 is practically similar in July 1956, a period of moderate solar activity. These features are analogous to those observed at the other low-latitude stations during local winter. However, the large diurnal variation in f_oF_2 despite the very small variation in the intensity of solar radiation during a winter day at such high latitudes seems to be an abnormal feature.

Figure 5 shows the diurnal variation of f_oF_2

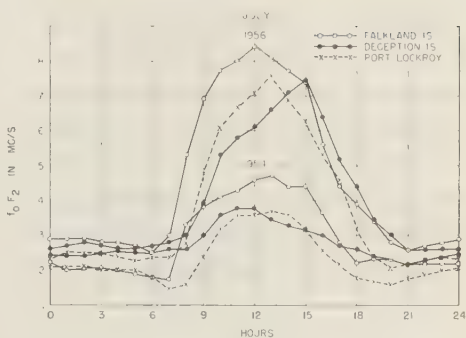


Fig. 4. The diurnal variations of the critical frequency of the F_2 layer in a winter month (July) at South American stations. Note that there is no difference in the nature of variations at these stations.

at these stations during a summer month (November) of 1954 and 1956. During the summer of 1954, the f_oF_2 at Falkland attains maximum values at noon and at late evening hours. At a more southerly station, Deception Island, there is hardly any diurnal variation of f_oF_2 . Still more southerly, at Port Lockroy, the maximum of f_oF_2 is observed at midnight and the minimum at midday. During the high-sunspot year 1956, the midday maximum of f_oF_2 at Falkland is more pronounced, a weak maximum of f_oF_2 is observed during midday hours at Deception, and the variation of f_oF_2 at Port

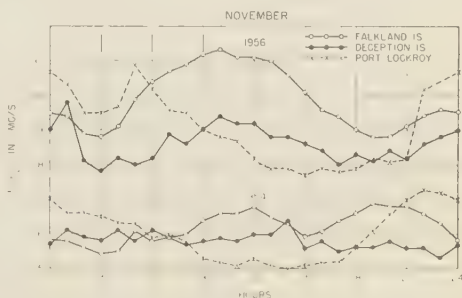


Fig. 5. The diurnal variations of the critical frequency of the F_2 layer in a summer month (November) at South American stations. Note the difference in the variations at these stations. Falkland shows a maximum but Port Lockroy a minimum of f_oF_2 around the noon hours, and Deception shows the least variation of f_oF_2 during the day.

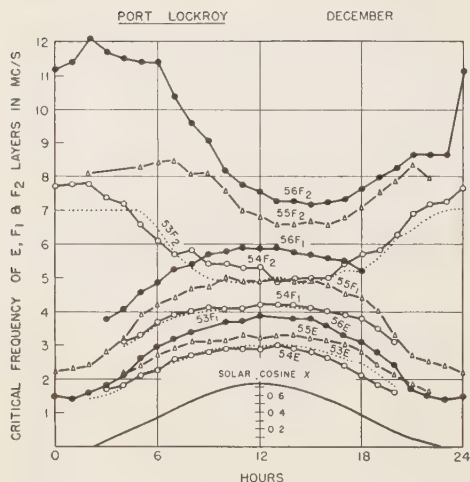


Fig. 6. The diurnal variations of the critical frequencies of the E , F_1 and F_2 layers at Port Lockroy during a summer month of the years 1953 to 1956. Note the regular variations of f_oE and f_oF_1 but the abnormal variation of f_oF_2 with a minimum at midday in all the years.

Lockroy is enhanced with the maximum still at midnight and the minimum at midday.

Port Lockroy, thus, experiences an abnormal variation of f_oF_2 during the local summer months; the change in the type of the diurnal variation occurs somewhere between Deception and Port Lockroy.

A comparison is made of the diurnal varia-

tions of the critical frequencies of all other layers of the ionosphere at Port Lockroy during a summer month. Figure 6 shows the diurnal variations of f_oE , f_oF_1 , f_oF_2 at Port Lockroy during the month of December of the years 1953 to 1956. The variation of the cosine of the zenith angle of the sun is also shown in the figure for a comparison between the variation of critical frequencies and the altitude of the sun. The critical frequencies of the E and F_1 layers show a maximum at noon and follow faithfully the progress of the solar altitude. The layers do not die out after sunset, owing to the extreme shortness of the night and because the sun is only a few degrees below the ground horizon even at midnight. The critical frequencies of the F_2 layer show a minimum during the afternoon hours and a maximum at midnight in any of the years. The diurnal variation is more pronounced during a high-sunspot year. The average frequency of any of the layers increases progressively from 1954 to 1956 as the result of the increasing solar activity during the period. Long-time variations of the intensity of the solar radiation has definite influence on the ionization of the ionospheric layers at Port Lockroy. *Bellchambeau and Piggott* [1958] have reported similar abnormal diurnal variations of f_oF_2 at Halley Bay (75°S, 27°W) during the summer month of December 1957.

Comparison of the F_2 layer at southern stations in east and west zones. In order to a

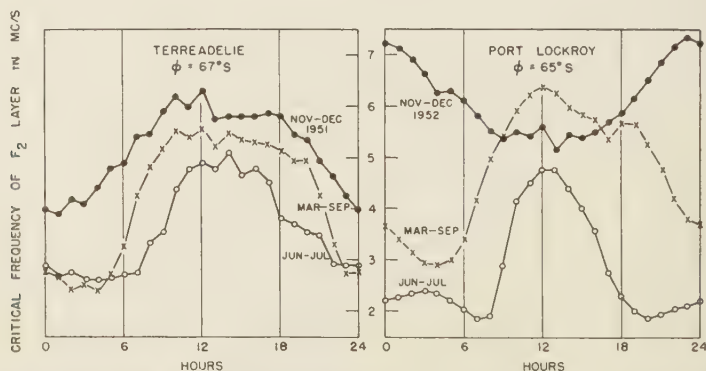


Fig. 7. Diurnal variations of the critical frequency of the F_2 layer in the three seasons at Terrea Delie and Port Lockroy. Note that the variations are similar at the two stations in winter (June-July) and in equinoxes (March-September) but opposite in the summer (November-December).

tain whether the abnormal diurnal variation of f_oF_2 is characteristic of that zone only, and diurnal variations of f_oF_2 at stations in different zones are compared. Figure 7 shows the diurnal variations of f_oF_2 during the different seasons at Port Lockroy (geographic latitude $\approx 35^\circ\text{S}$) in the west zone and at Terre Adelie (geographic latitude $= 67^\circ\text{S}$) in the east zone. At Terre Adelie, f_oF_2 attains the maximum value at midday and minimum at midnight in all seasons, a feature characteristic of the F_2 layer at high latitudes. At Port Lockroy the variation of f_oF_2 is similar to that at Terre Adelie during the July and March–September, though the range of variation is larger at Port Lockroy. During November–December the diurnal variations of f_oF_2 at the two stations are opposite to each other. This shows that the abnormal diurnal variation of f_oF_2 during summer months is characteristic of the west zone only.

In Figure 8 are shown the contour maps of f_oF_2 in terms of hour of the day and month of the year at the stations Port Lockroy and Macquarie Island for 1953–1954, one of the quietest periods of the sun. Macquarie Island is in the east zone, where no abnormal diurnal variation of f_oF_2 is observed during the summer months. In these contour maps the abnormality of the F_2 region at Port Lockroy during the course of the day can be seen. The contour map for

Macquarie Island is much simpler, indicating the maximum ionization at the afternoon hours of December and the minimum at the after-midnight hours of June. All the contour lines are parallel to the sunrise and sunset lines. Thus the critical frequency of the F_2 layer at Macquarie Island for any of the hours is maximum in summer and minimum in winter. This corresponds to the similar behavior of the critical frequencies of the lower E and F_1 layers. The contour map of f_oF_2 for Port Lockroy is comparatively complicated. The maximum of daytime f_oF_2 is observed in April and October very much as in other middle-latitude stations. The night-time frequency does not follow the sun's altitude, and the contour lines are not parallel to sunset or sunrise lines. The sunrise effect on the variation of f_oF_2 is not much in evidence. The critical frequency of the F_2 layer is sensitive to the zenith angle of the sun as well as to a parameter depending on the position of the station. The former is prominent in the diagram for Macquarie Island whereas the latter seems to be more effective for Port Lockroy during summer months.

Discussion. One of the common explanations for the abnormal variations of f_oF_2 at some stations has been on the basis of contraction and expansion of the F_2 region. Ranzi [1957] attributed the midday decrease of f_oF_2

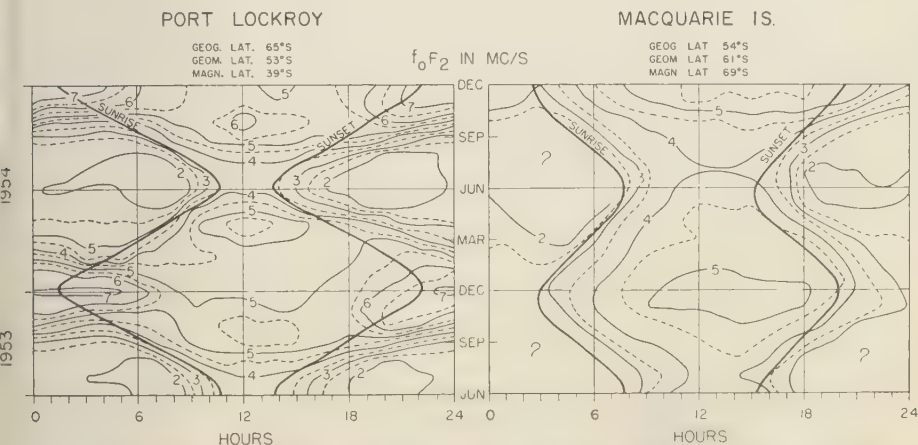


Fig. 8. Contour maps of the critical frequency of the F_2 layer as a function of hour and month at Port Lockroy and Macquarie Island. Note the regular variation of f_oF_2 at Macquarie.

at Deception Island to the vertical drift, which he suggested was very pronounced in that zone. In Figure 9 are shown the diurnal variations of the height of maximum ionization ($h_m F_2$), semithickness (TF_2), height of the base ($h_o F_2 = h_m F_2 - TF_2$), maximum electron density ($N_m F_2 = 1.24 \times 10^{-4} f_o^2 F_2$), and the total electron content in a vertical column up to h_m ($n F_2 = 2/3 TF_2 \times N_m F_2$) of the F_2 layer at Port Lockroy during the summer months of December 1953 and January 1954, based on the ionospheric characteristics published by Radio Research Station, Slough. The height of the base or of the maximum electron concentration does not change appreciably with the time of the day. The semithickness of the F_2 layer remains fairly constant at about 80 km. Further, owing to the constant thickness of the layer, the maximum electron concentration and the total electron content show similar diurnal variations, namely a maximum at midnight and a minimum at midday. These facts do not support any hypothesis to explain the decrease of the $f_o F_2$ in terms of the thermal expansion or vertical drifts in the F_2 region.

Rastogi [1959b] has indicated that the amount of ionization in the F_2 region at any place on the earth is made up of the ionization produced at the place by the solar radiations

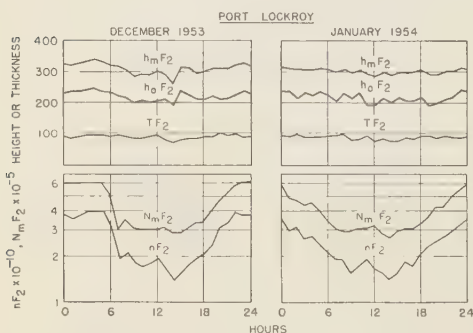


Fig. 9. Diurnal variations of the height of maximum electron density ($h_m F_2$), the height of the base ($h_o F_2$), semithickness (TF_2), the maximum electron density ($N_m F_2$), and the total electron content ($n F_2$) of the F_2 layer at Port Lockroy during the summer months. Note the minimum values of $N_m F_2$ or $n F_2$ during the noon hours and almost constant thickness of the F_2 layer during the whole day.

plus the ionization transported from neighboring regions. Some of the abnormalities of the F_2 region may be explained by assuming the horizontal transport of ionization in the F_2 region along the meridian in the direction away from the equator in the forenoon and toward the equator in the afternoon. Such a meridional motion of ionized mass would be greatly affected by the earth's magnetic field. The transport of ionizations in the forenoon hours from a higher F_2 region at the equator to a comparatively lower F_2 region at greater latitudes would be along the lines of force of the earth's magnetic field. Here the above hypothesis is extended by assuming that the direction of motion of the ionization from the equator would be parallel to the direction of maximum change of the magnetic dip, i.e. the gradient of the magnetic dip. Figure 10 shows the world map with the isomagnetic dip lines. The lines of the gradient of magnetic dip are drawn starting from each 10° of longitude on the geographic equator and such that they are perpendicular to the dip line at any point. These lines at any point represent the shortest distance between two isodip lines and hence the direction of flow of ionization along the lines of force. The isodip lines are highly curved in the South American region, but almost parallel to the geographic latitude lines in the Asian and Australian zones. Consequently, the dip gradient lines are nearly parallel to meridians in the Asian zone but depart appreciably from them in the South American zone. The transport of ionization away from the equator would be guided almost along the geographic meridian in the Asian zone but not in the South American zone. This would affect the diurnal variations of $f_o F_2$ at different places. Further, from Figure 10 the comparative rarity of the gradient lines in the southern west zone can be noted. If we assume, qualitatively at least, the amount of transported ionization to be proportional to the density of gradient lines in that zone, it is easily understood that Port Lockroy and Deception would not receive much of the transported ionization from the equator during the daytime hours. This explains the abnormally low F_2 layer ionization at the South American stations during the daytime.

Summary. 1 In the west zone the ionization

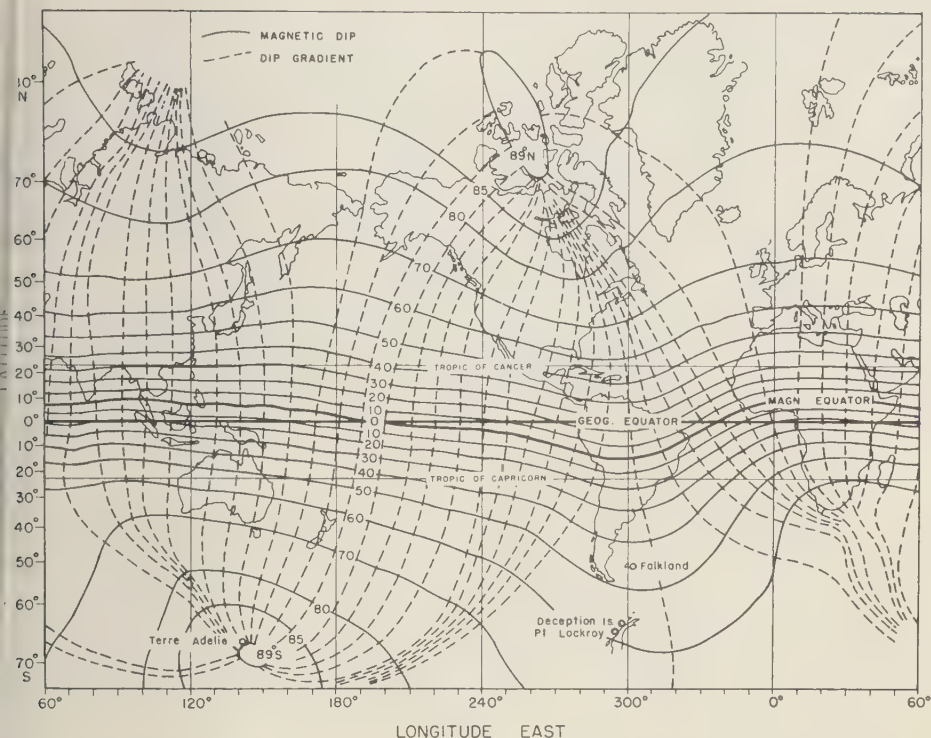


Fig. 10. World map showing the lines of equal dip and the lines of the gradient of dip. Note the comparative rarity of gradient lines in the South American region.

on in the F_2 region of the ionosphere at a station in the southern hemisphere is lower than that at a station of equal magnetic dip in the northern hemisphere. No such discrepancy is noticed in the east zone.

2. The distribution of midnight f_oF_2 is controlled more by the magnetic latitude than by the geographic or geomagnetic latitude, especially around the December solstice.

3. Abnormally high values of f_oF_2 are observed at southern stations of the west zone during the midnight hours of local summer months.

4. During a summer day, the variation of E and f_oF_1 at Port Lockroy is similar to that of the solar altitude, but the variation of f_oF_2 is opposite to that of $\cos \chi$; the value of f_oF_2 reaches a maximum at midnight and a minimum around noon.

5. Such a peculiar variation of f_oF_2 is characteristic of the southern west zone only and is not observed at other high-latitude stations in either the south or the north.

6. The midday depression of f_oF_2 during the summer day at Port Lockroy cannot be accounted for by the vertical drifts of ionization.

Conclusions. The elements of the earth's magnetic field as measured at the ground may have pronounced effect on the behavior of the F_2 region of the ionosphere. It is hoped that the ionospheric data obtained at the Antarctic ionospheric stations established during the IGY will be helpful in studying the relation between the ionosphere and geomagnetism in that region of the earth.

Acknowledgments. The author is indebted to Professor K. R. Ramanathan for his keen interest and stimulating discussions. He is also grateful to

the members of the Upper Atmosphere Section of the National Research Council of Canada for valuable suggestions during the preparation of the article.

REFERENCES

- Bellchambers, W. H., and W. R. Piggott, Ionospheric measurements made at Halley Bay, *Nature*, 182, 1596-1597, 1958.
- Lepechinsky, D., and G. Pillet, Contribution à l'étude de l'influence du magnetisme terrestre sur l'ionisation de la couche F_2 , *Notes préliminaires du Laboratoire National de Radio-électricité*, 139, 1950.
- Ranzi, I., Report of the Italian National Committee, *Proc. XIIIth General Assembly of URSI*, Boulder, vol. XI, pt. 3, p. 59, 1957.
- Rastogi, R. G., Magnetic control on the variation of the critical frequency of the F_2 layer of the ionosphere, *Can. J. Phys.*, 37, 874-879, 1959a.
- Rastogi, R. G., The diurnal development of the anomalous equatorial belt in the F_2 region of the ionosphere, *J. Geophys. Research*, 64, 727-732, 1959b.
- Shapley, A. H., Clues to ionospheric conditions in the southern auroral zone, Geophysical monograph 1, *Antarctica and the IGY*, pp. 86-100, 1956.

(Manuscript received September 14, 1959;
revised December 4, 1959.)

World Maps of F_2 Critical Frequencies and Maximum Usable Frequency Factors for Use in Making Ionospheric Radio Predictions

DONALD H. ZACHARISEN AND VAUGHN AGY

*National Bureau of Standards
Boulder, Colorado*

Abstract. A recent publication of the National Bureau of Standards is described which gives charts for use in predicting the classical F_2 -layer maximum usable frequency for any sunspot number. The charts present f_oF_2 and the M4000 factor at sunspot number 50, and the rates of change of these parameters with sunspot number. A short history of NBS activity in this field is presented, together with a comprehensive list of publications giving methods and tools used in making predictions by various organizations throughout the world.

The purpose of this note is to call attention to a new presentation of ionospheric data for estimation of the classical F_2 -layer MUF for any sunspot number [Zacharisen, 1959]. The work has been published in the form of an NBS *Technical Note* which is on sale at the United States Department of Commerce, Office of Technical Services, Washington 25, D. C.

Almost every major country in the world has found it advantageous to prepare and publish maps, charts, or tables that can be used as a basis for selecting frequencies for sky-wave radio transmission between two points on the surface of the earth. (A list of the laboratories that provide prediction services is appended.) Interest in this field arose in the late 1930's and was accelerated during World War II.

The number of ionospheric sounding stations furnishing the data on which these predictions are based was less than 10 in the 1930's. By 1957, however, there were approximately 80, and this number was doubled during the IGY. Nevertheless, one of the major sources of error in making MUF predictions is the lack of a sufficient number of ionosonde stations. A relatively poor geographical distribution of existing stations also contributes to the error. They are concentrated in the more populous areas of the continents: there are very few in the Pacific Ocean and fewer still in the Atlantic Ocean; so there are many more stations in the northern than in the southern hemisphere.

Monthly predictions were first published in

1939 [Gilliland, Kirby, and Smith, 1939] by the National Bureau of Standards, consisting of one chart in which the F_2 MUF was plotted as a function of time of day for various distances of transmission. Publication of the predictions was interrupted in 1941 when the United States entered the war.

In 1942, the monthly prediction for the armed services took the form of contours of MUF on a grid of latitude vs. local time, for a distance of 2200 miles. A table of factors was given to obtain the MUF for shorter distances. In 1943, as a result of the receipt of ionospheric data from the far east, which indicated a longitudinal difference in F_2 -layer conditions, two contour charts of F_2 -layer MUF were issued, one for 75° west longitude and the other for 100° east longitude. The MUF for other longitudes was obtained by linear interpolation between these charts. It was early recognized that this longitude variation was an effect of the earth's magnetic field, and in 1944, as a result of an International Radio Conference held in Washington, D. C., the three-zone system based on the magnetic field was adopted. The same zones are still used in the CRPL-D Series 'Basic Radio Propagation Predictions,' although in 1958 separate charts for the two parts of the intermediate (I) zone were initiated.

The approximately linear relationship between the critical frequency and the sunspot number has long been recognized and was used in all the prediction methods described above.

All these predictions were based upon a particular sunspot number, estimated in advance. The charts, therefore, were of transient usefulness and had to be completely redrawn for each succeeding month.

The latest effort, resulting in *NBS Technical Note 2*, enables the user to estimate the F_2 -layer MUF for any sunspot number. In this presentation world maps for particular hours of Universal Time are used instead of the zone system. Up to 17 years of data were utilized. The values used in preparing the prediction charts are based on the observed regression of the F_2 -layer ordinary-wave critical frequency (f_oF_2) and the F_2 -layer MUF factor for a transmission distance of 4000 km (M4000 factor) on the 12-month-running-average Zurich sunspot number (RASSN).

The four ionospheric parameters used are: monthly median f_oF_2 at RASSN 50, the slope of the regression line of f_oF_2 on RASSN, the monthly median M4000 factor at RASSN 50, and the slope of the regression line of M4000 on RASSN. The first three of these are presented in the form of world contour maps for each even hour of Universal (Greenwich) Time; the fourth is given in the form of contours plotted against geomagnetic latitude and local time. It was decided from the results of a pilot study [Crow and Zacharisen, 1959] that it was unnecessary to draw maps of the fourth parameter in more detail because of the relatively large percentage error in its estimation from sample data. Charts are given for January, March, June, July, September, and December. Charts for the remaining months of the year are in preparation for publication.

Also included in the *Technical Note* is a table of gyrofrequencies for converting f_oF_2 to f_zF_2 and a nomogram for interpolating between f_zF_2 and the 4000-km MUF, the latter being obtained by multiplying f_oF_2 by M4000. Thus, given the RASSN (the charts are valid up to about RASSN 150), the classical F_2 -layer MUF may be estimated for any path, month, and time of day.

Other publications of this nature which include the sunspot number as an independent parameter are the tables of the Radio Physics Laboratory of Canada and the charts of the Radio Research Laboratories of Japan. The

Canadian tables are restricted to latitudes greater than 35° north. The Japanese charts consist of world contour maps of f_oF_2 and 4000-MUF at sunspot numbers 10 and 15. Values at other sunspot numbers are obtained by linear interpolation (or extrapolation). The charts of F_2 -4000-MUF are based on the assumption that the ratio of F_2 -4000-MUF to f_oF_2 (i.e., the M4000 factor) is independent of sunspot number.

REFERENCES

- Crow, E. L., and D. H. Zacharisen, The error in the prediction of F_2 maximum usable frequency from world maps based on sunspot number, in *Symposium on Statistical Methods in Ionospheric Wave Propagation* (U.C.L.A. June 18-20, 1959), edited by W. C. Hoffman, Pergamon Press, New York, 1959.
- Gilliland, T. R., S. S. Kirby, and N. Smith, Characteristics of the ionosphere at Washington, D. C., January 1939, *Proc. IRE*, **27**, 226 (1939).
- Zacharisen, D. H., World maps of F_2 critical frequencies and maximum usable frequency factors, *NBS Tech. Note 2*, Office of Technical Services, U. S. Department of Commerce, April 1959.
- LABORATORIES PROVIDING PREDICTION SERVICES
- Central Radio Propagation Laboratories (CRPL), Basic radio propagation predictions, series A, National Bureau of Standards, U.S.A.
- Centre National d'Etudes des Télécommunications, Prévisions pour la propagation radioélectrique SPIN-A-I, SPIN-A-II, and SPIN-A-III, Section de prévisions Ionosphériques Nationales, Saclay (Seine and Oise), France.
- Department of Scientific and Industrial Research, Predictions of radio wave propagation conditions, Radio Research Station, Slough, England.
- Diretoria de Electronica da Marinha, Previsões ionosféricas, MUF/RIO, Ministerio da Marinha, Ilha das Cobras, Rio de Janeiro.
- Fernmeldetechnisches Zentralamt (FTZ), Darmstadt, Federal German Republic.
- Ionospheric Prediction Service, Ionospheric predictions, series A, B, H, P, R, S, and W, Department of the Interior, Commonwealth of Australia.
- Instituto Geofísico de Huancayo, Boletín de propagación, Ministerio de Fomento, Huanca, Peru.
- Ministerio de Marina, Predicciones de radiopropagación ionosférica, Dirección General del Material de Comunicaciones Navales Laboratorios Ionosféricos Armada Republica Argentina, Buenos Aires, República Argentina.
- Scientific Research Institute of Terrestrial

ism (NISMIR) Ionospheric and Radio-wave Propagation, Ministry of Communications, Moscow, U.S.S.R.

o Physics Laboratory Report 1-1-3, Prediction of optimum traffic frequencies for northern latitudes, Defence Research Board, Canada, November 1954.

o Research Committee, Ionospheric predictions, Council of Scientific and Industrial Research, New Delhi, India.

Radio Research Laboratories, World maps of

F_2 critical frequencies and maximum usable frequencies for 4000 km, Ministry of Posts and Telecommunications, Tokyo, Japan, August 1958.

South African Council for Scientific and Industrial Research, Radio propagation predictions for Southern Africa, Telecommunications Research Laboratory, Johannesburg, Union of South Africa.

(Manuscript received November 30, 1959.)

On the Propagation of ELF Radio Waves and the Influence of a Nonhomogeneous Ionosphere

JAMES R. WAIT

National Bureau of Standards
Boulder, Colorado

Abstract. The model assumed consists of a spherical earth surrounded by a concentric ionosphere whose electron density increases exponentially with height. This elaboration of the usual homogeneous model appears to explain the observed attenuation for terrestrial propagation as a function of frequency in the range 100 c/s to 1 kc/s.

The propagation of electromagnetic waves at frequencies of the order of 1 kc/s and less has been studied on several occasions recently [Capman and Macario, 1956; Lieberman, 1957]. At these extremely low frequencies (denoted ELF) it appears that the electromagnetic energy is channeled between the earth and the ionosphere in a wave guide. In this note a theoretical model is introduced which takes into account the increase of the electron density with height in the ionosphere. This elaboration of the usual homogeneous model is apparently necessary to explain the observed attenuation as a function of frequency in the range 100 c/s to 1 kc/s. The need for such a modification of the theory has recently been pointed out by E. T. Pierce (private communication).

The assumed theoretical model chosen is a homogeneous conducting spherical earth of radius a surrounded by a concentric conducting ionospheric shell of inner radius $a + h$. It is convenient to introduce the usual spherical coordinate system (r, θ, ϕ) . Using this model and assuming a vertical electric dipole source located at $\theta = 0, r = a$, an expression for the vertical electric field is the following [Wait, 1957]:

$$E_r = \frac{I ds \eta_0 \nu(\nu + 1)}{4kha^2 \sin \nu\pi} P_\nu(-\cos \theta) \quad (1)$$

where

I = average current in the source dipole.

ds = length of source dipole.

η_0 = intrinsic impedance of air space $\cong 120\pi$ ohms.

$k = 2\pi/\text{wavelength}$.

$P_\nu(-\cos \theta)$ = Legendre function of argument $-\cos \theta$ and (complex) order ν .

The quantity ν is to be found from the boundary conditions of the problem. For present purposes it is adequate to use the following result [Wait, 1957, 1960]:

$$\nu + \frac{1}{2} \cong kaS_0 \quad (2a)$$

with

$$S_0 = 1 - \frac{i}{2kh} \left(\frac{Q}{\bar{N}} + \frac{1}{N_0} \right) \quad (2b)$$

where N_0 and \bar{N} are the complex refractive indices of the earth and the lower edge of the ionosphere, respectively, and Q is a complex factor which takes into account the stratification in the ionosphere. (This factor is discussed below.)

Now, at ELF, $|N_0| \gg |\bar{N}|$, and thus

$$S_0 \cong 1 - \frac{i^{3/2} Q}{2(\sigma_i \mu_0 \omega)^{1/2} h} \quad (3)$$

in terms of the effective conductivity of the ionosphere. If $|\nu| \gg 1$ and θ is not near 0 or π it is permissible to employ the asymptotic approximation

$$P_\nu(-\cos \theta) \cong \left(\frac{2}{\pi \nu \sin \theta} \right)^{1/2} \cdot \cos \left[\left(\nu + \frac{1}{2} \right) (\pi - \theta) - \frac{\pi}{4} \right] \quad (4)$$

Thus, apart from a constant factor, the field E_r may be identified as the linear combination of

two peripheral waves of the form (for $\theta < \pi$)

$$\frac{1}{(\sin \theta)^{1/2}} e^{-ika S_0 \theta}$$

and

$$\frac{1}{(\sin \theta)^{1/2}} e^{-ika S_0 (2\pi - \theta)} e^{i\pi/2}$$

The attenuation of these waves in nepers per unit length is given by the quantity $-Im(kS_0)$. In the case of a homogeneous ionosphere the factor Q is replaced by unity and the attenuation thus varies as $f^{1/2}$. This is the case considered at great length by *Schumann* [1954]. Experimental data of attenuation rates in the range 100 c/s to 1 kc/s indicate that the variation is more like f^γ , where γ is less than but near 1 [Chapman and Macario, 1956; R. E. Holzer, private communication]. To obtain a theoretical behavior more in accord with data it is desirable to introduce an ionospheric model whose refractive index is increasing with height. One approach is to consider a layered model. In fact, it was shown previously [Wait, 1958] that a two-layer model will reconcile certain apparent discrepancies in the relative attenuation rates at very low frequencies (VLF) and extremely low frequencies (ELF).

The other approach, adopted here, is to let the refractive index increase or decrease from some initial value \bar{N} at height h in a monotonic fashion. Choosing an exponential variation, the refractive index as a function of height is explicitly

$$N(z) = 1.0 \quad \text{for } 0 < z < h \\ = \bar{N} \exp [(z - h)/l] \quad \text{for } z > h$$

where l is a scale factor. If $l > 0$, the refractive index is increasing with height; if $l < 0$, the refractive index is decreasing with height. See Figure 1. It has been shown elsewhere [Wait, 1958] that the factor Q has the forms

$$Q = \frac{K_0(i\bar{N}kl)}{K_1(i\bar{N}kl)} \quad \text{if } l > 0 \quad (5a)$$

and

$$Q = \frac{I_0(-i\bar{N}kl)}{I_1(-i\bar{N}kl)} \quad \text{if } l < 0 \quad (5b)$$

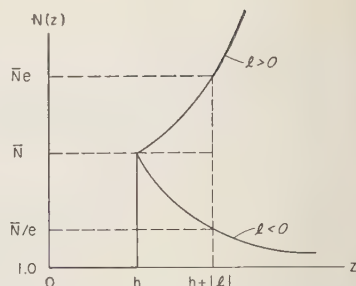


Fig. 1. Magnitude of refractive index as a function of height for exponential profiles.

where I_0 , I_1 , K_0 , and K_1 are modified Bessel functions with their conventional meaning.

The preceding results are valid for $|\bar{N}| \gg 1$ which is well satisfied at ELF in the region where these waves are reflected. To the same approximation

$$\bar{N} \cong \left[1 - i \frac{\bar{\omega}_r}{\omega} \right]^{1/2} \cong \left(\frac{\bar{\omega}_r}{\omega} \right)^{1/2} e^{-i\pi/4} \quad (6)$$

where $\bar{\omega}_r = \bar{\omega}_0^2/\bar{\nu}$ in terms of the plasma frequency $\bar{\omega}_0$ and collision frequency $\bar{\nu}$. It then follows that the imaginary and real parts of the propagation constant kS_0 are given by

$$Im kS_0 = -\frac{1}{2h} \left(\frac{\omega}{2\omega_r} \right)^{1/2} P(x) \quad (7a)$$

and

$$Re kS_0 = k + \frac{1}{2h} \left(\frac{\omega}{2\omega_r} \right)^{1/2} P'(x) \quad (7b)$$

where

$$P(x) = |Q| \sin \left(\frac{\pi}{4} - \arg Q \right) \sqrt{2} \quad (8a)$$

$$P'(x) = |Q| \cos \left(\frac{\pi}{4} - \arg Q \right) \sqrt{2} \quad (8b)$$

and where

$$x = |kl\bar{N}| = |kl| (\omega_r/\omega)^{1/2}$$

The functions $P(x)$ and $P'(x)$ are plotted in Figures 2a and 2b for the cases $l > 0$ and $l < 0$ respectively. All these quantities approach unity if x is sufficiently large. This limiting case corresponds to the homogeneous ionosphere. The $P(x)$ is the modification of the attenuation and

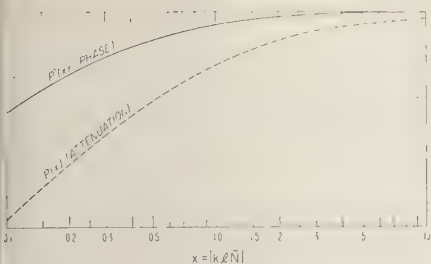


Fig. 2a. Modification of attenuation and phase for an exponential profile ($l > 0$).

$P(l)$ is the modification of the phase resulting from a nonhomogeneous ionosphere.

It may be observed from the curves in Figure 2 that the attenuation is generally lower if the refractive index is an increasing function with height. On the other hand, as seen from Figure 2 the attenuation would be increased where the refractive index is decreasing with height.

To illustrate the behavior of the attenuation curves as a function of frequency they are plotted in Figure 3 in terms of decibels per 1000 km of path length for a frequency scale from 50 c/s to 15 kc/s. Values of the scale distance l are indicated in kilometers. The height h of the lower edge of the ionosphere is taken to be 90 km. The value chosen for $\bar{\omega}_r$ is 10^5 , which is not inconsistent with previous work.

It is seen that the curve for $l = \infty$ corresponds

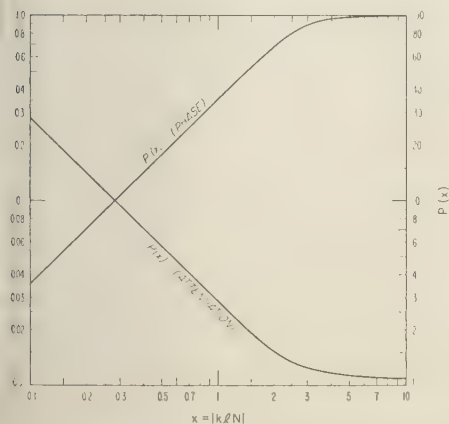


Fig. 2b. Modification of attenuation and phase for an exponential profile ($l < 0$).

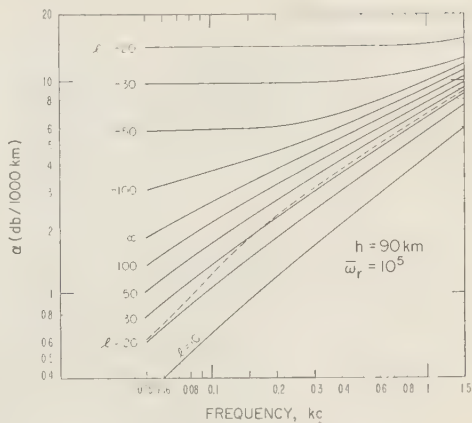


Fig. 3. Attenuation at ELF for an exponential profile. Solid curves, calculated; dashed curve, measured [Chapman and Macario, 1956].

ing to a homogeneous ionosphere has a slope of $1/2$, as it should. Depending on the sign of l and its magnitude, the slope may be modified considerably for the inhomogeneous ionosphere. The dotted curve in Figure 3 corresponds to night-time experimental data from Chapman and Macario [1956], who obtained it from the spectral analyses of large numbers of atmospherics recorded in London. Clearly this experimental curve appears to fit fairly well on the $l = 30$ km curve, at least in the range 100 c/s to 1.5 kc/s. There is no reason to expect any better fit than this because of the idealized profile assumed. Furthermore, the atmospherics analyzed by Chapman and Macario often originated quite near the receiving location in terms of a wavelength. For example, at 100 c/s the wavelength is 3000 km, and for distances less than this it is not strictly permissible to replace the Legendre function by only the first term in its asymptotic expansion. When using vertical electric dipole-type antennas it may be shown that, at distances small compared with a wavelength but large compared with the ionospheric heights, the vertical field varies as $\log a\theta$ rather than as $(a\theta)^{-1/2}$ [Wait, 1960]. The consequence is that the apparent attenuation is abnormally small at the lower end of the frequency range. However, in considering the amplitude of the vertical electric field, this effect is relatively minor in the frequency range above 100 c/s if the dis-

tance to the source of the atmospheric is greater than 1000 km.

In the preceding discussion, the influence of the earth's magnetic field has tacitly been neglected. This influence may be important¹, but until more data become available it hardly seems worthwhile to make any further elaborations to the model.

Acknowledgment. I should like to thank Dr. E. T. Pierce for his comments on this subject and Mrs. Anabeth Murphy for assistance with the calculations.

¹ It may be demonstrated [Wait, 1960] that for ELF

$$\frac{\text{Attenuation with magnetic field}}{\text{Attenuation without magnetic field}}$$

$$\simeq \frac{\cos(\tau/2)}{(\cos \tau)^{1/2}} \geq 1$$

where $\tan \tau =$

$$\frac{\text{Vertical component of gyro-frequency}}{\text{Collision frequency}}.$$

REFERENCES

- Chapman, F. W., and R. C. V. Macario, Propagation of audio frequency radio waves to great distances, *Nature*, 177, 930, 1956.
- Lieberman, L., Anomalous propagation below 500 c/s, *Proc. Symposium on Propagation of VLF Radio Waves*, 3, paper 25, Boulder, Colorado, 1957.
- Schumann, W. O., "Über die Strahlung langwelliger Wellen des horizontalen Dipols in dem Lufthohlraum zwischen Erde und Ionosphäre," *angew. Phys.*, 6, 225, 1954.
- Wait, J. R., On the mode theory of VLF ionospheric propagation, *Rev. Geofis. pura e appl.*, 37, 103, 1957. (Paper presented at International URSI Conference of Radio Wave Propagation, Paris, September 1956.)
- Wait, J. R., An extension to the mode theory of VLF ionospheric propagation, *J. Geophys. Res.*, 63, 125, 1958.
- Wait, J. R., Terrestrial propagation of VLF radio waves—a theoretical investigation, *J. Research NBS*, 64D, 153–203, 1960.

(Manuscript received November 7, 1959; revised November 23, 1959.)

Night-Time Equatorial Propagation at 50 Mc/s First Results from an IGY Amateur Observing Program

M. P. SOUTHWORTH

*Radio Propagation Laboratory
Stanford University
Stanford, California*

Abstract. During IGY the American Radio Relay League collected radio amateur reports of 50- and 144-Mc/s ionospheric propagation, evaluated them, and transcribed them onto punched cards. Analysis of 50-Mc/s equatorial intercepts, begun this summer at Stanford University, has revealed three apparently related types of nocturnal, low-latitude propagation: (1) long-range transequatorial, as first noticed by amateurs in 1947; (2) medium-range, between stations making transequatorial contacts and stations near the magnetic equator; (3) short-range, similar to sporadic *E* but occurring regularly with the first two types. Where ever they appear these modes are present almost nightly during certain months, and evening propagation of frequencies up to 1.5 times the maximum daylight MUF is not uncommon. Comparison of transequatorial results in the Americas and the Far East has shown that seasonal behavior is not the same at all meridians. Pronounced negative correlation with magnetic activity is a world-wide feature, however, which suggests a direct relation to equatorial spread *F*. Quantitative professional data at the frequencies of interest are rather rare, but comparisons with Professor Dueño's backscatter soundings made at the University of Puerto Rico indicate that the 20- and 40-Mc/s transequatorial propagation seen there is not the same as what amateurs experience on 50 to 75 Mc/s.

Introduction. Radio amateurs have shown many years that 50-Mc/s signals can conveniently cross the magnetic equator during evening hours. The first contacts of this type, between Mexico and Argentina, were made in August 1947 [Tilton, 1947]. *QST*, the amateur journal, reported much more so-called transequatorial scatter (referred to as TE in this paper) during the following 4 years. Ferrell [1951], describing the RASO amateur project, told URSI about this new mode at its April 1951 meeting. Little or no TE was observed throughout the sunspot minimum, but it reappeared in 1955 [Tilton, 1955]. Amateur experience since January 1957 is especially well documented, and the American Radio Relay League's IGY Propagation Research Project (PRP) [Southworth, 1959] has provided us with readable data.

In this latest program several hundred amateurs supplied semimonthly reports of all propagation observed on 50-54 and 144-148 Mc/s that might involve the ionosphere. The logs were carefully screened at a special office established by the author while with ARRL. A

punched card containing the following information was prepared for each usable item: frequency band, date, begin and end times of intercept or watch period, observer call sign, observer coordinates, other station call sign, other station coordinates, antenna heading for best results, signal strength and fading characteristics, and type of propagation and type of report (two-way, heard or negative). More space on each card was available for computed information such as path length and mid-point coordinates. These cards and the original logs are the basis of most of the content of this paper.

There have been few professional TE experiments at VHF. Villard, Stein, and Yeh [1957] have reported on 23-, 30-, 40-, and 46-Mc/s backscatter soundings made in the Virgin Islands during August and September 1956. Southern echoes having ranges of 5500 to 11,000 km were seen often between 1500 and 2100 hours local time. Similar soundings on 20 and 40 Mc/s have been made by Dueño [1959] at the University of Puerto Rico since the spring of 1957. He observes the long TE echoes and also a class

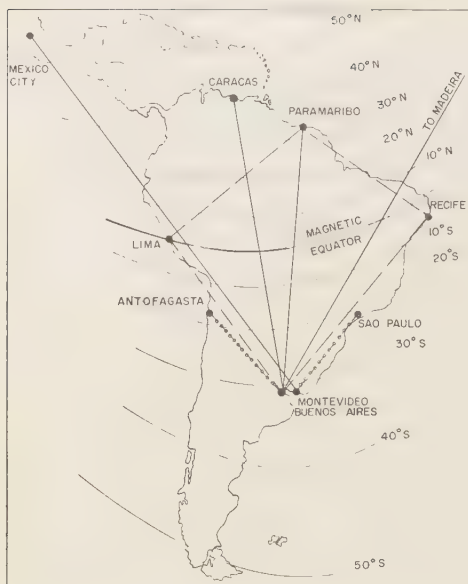


Fig. 1. Typical evening 50-Mc/s paths observed by radio amateurs in South America.

of medium-range returns from 3000 to 4000 km (the vicinity of the magnetic equator). These last tend to appear as the TE echoes fade out. The several 50-Mc/s circuits set up by Bowles and Cohen in South America [Gates, 1959] included a 2580-km transequatorial path. They found a kind of forward scatter which is coincident with spread *F* above the mid-point. The anomalous evening signals studied by CRPL [Bateman and others, 1959] between Okinawa and the Philippines at 50 Mc/s, although not transequatorial, have similar characteristics.

Typical evening paths. Nocturnal 50-Mc/s propagation reported from South America falls in three main categories on the basis of path length. Figure 1 shows typical examples of each. The solid-line transequatorial paths are the longest, ranging from 4000 to nearly 9000 km. They are roughly bisected by the magnetic equator, and most end points have magnetic dip angles between 20° and 60° . Amateurs in the Buenos Aires–Montevideo area, for example, contact such places as Paramaribo, Caracas, and Mexico City almost nightly during several months of the year. Signals from the Madeira Islands are heard only a little less often.

Paths between 2500 and 4000 km in length have been classed as medium range; four examples show on the map as dashed lines. Lima–Recife stations are often heard in both Buenos Aires–Montevideo and Paramaribo. Usually the latter are also hearing each other and additional transequatorial stations at such times.

The third category, short range, includes intercepts between 200 and 2500 km. Dotted paths between Buenos Aires–Montevideo and Sao Paulo and Antofagasta are typical, though the distances here resemble single-hop sporadic *E*, 50-Mc/s signals are heard on these paths at the same times and with the same regularity as transequatorial ones.

The bulk of amateur activity in South America is in the cities along the coasts and in the north-central part of Argentina. There is much farther south than Buenos Aires and little in the central part of the continent, though amateur radio can provide data in more places than professional experiments, though there are still many possible circuits whose 50-Mc/s behavior is not known. For this reason, defining of three modes is more for convenience.

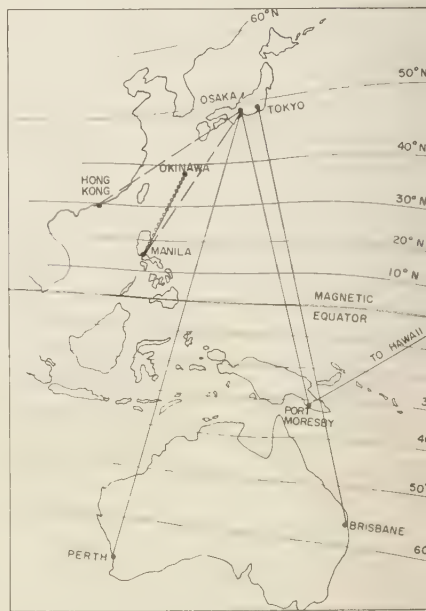


Fig. 2. Typical evening 50-Mc/s paths observed by radio amateurs in the Far East.

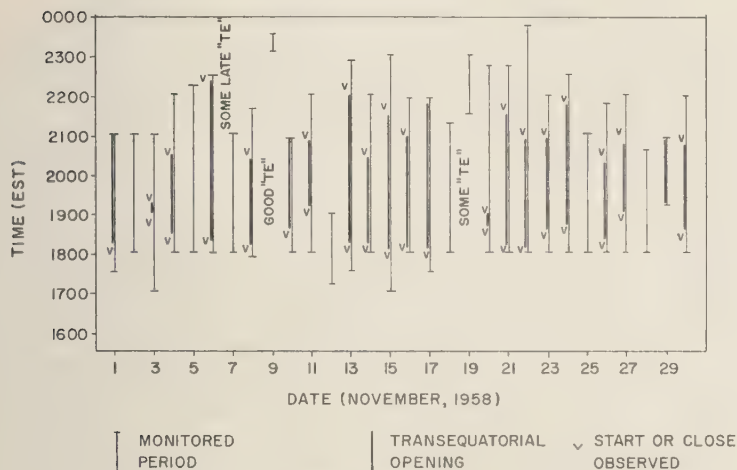


Fig. 3. Evening monitoring by station PZ1AE, Paramaribo, Surinam, during November 1958.

to state that they are separate and distinct. Short range merges into medium range, medium range into TE, and sometimes stations can communicate over a practically continuous range from a few hundred to 9000 km. Increased overseas 50-Mc/s activity since the solar maximum and the publicity given to equatorial scatter by PRP have combined to produce more data from other parts of the world. In March 1957, Australian amateurs heard their first evening 50-Mc/s signals from Japan. Regular work between southern Europe and southern Africa began in August of that year. Figure 2 shows typical paths in the Far East. Amateurs in Queensland and Papua Territory hear TE from Japan most frequently. Inception in West Australia and New South Wales is more usually regular F_2 . Papua to the Hawaiian Islands is a longer but quite consistent TE circuit. Two medium-range paths are shown up: Japan to Manila and Japan to Hong Kong. Apparently there are no amateurs at 50 Mc/s in the right places to reveal short-range evening propagation. CRPL observations of the Okinawa-Philippines circuit, however, show exactly the short-range characteristics one would expect from the amateur results in South America.

Usefulness of the amateur data. It may be asked, 'How reliable can your data be when so many radio hams just want to get on the air

and talk to each other?' 'Don't you have all those evening reports just because hams only operate between dinner and bedtime?' To begin with, 50-Mc/s amateur radio in low-latitude regions is not a casual affair. Long hours of listening are required, and several stations cooperate by making continuous beacon transmissions. In the case of nightly TE propagation, many amateurs are likely to be at their sets each evening waiting for it and to put off retiring as long as possible to catch the very end. PRP encouraged this attitude and took full advantage of it by requesting negative reports—dates and times when 50 Mc/s was checked to no avail—along with positive ones. Such reports can confirm activity any time there is a break in the records.

See Figure 3, which portrays the evening monitoring of PZ1AE, Paramaribo, Surinam, during November 1958. Light lines show when the operator was listening; heavy lines, when transequatorial signals were received. Even with 16 other stations located where they could detect South American TE, only three instances of it were discovered which PZ1AE did not report by himself. Furthermore, of the 21 evenings TE was observed at Paramaribo, the actual start and end times were pinpointed in 20 and 17 cases, respectively. These are reliable because PZ1AE, like many other amateurs, has trained himself to use existing commercial sig-

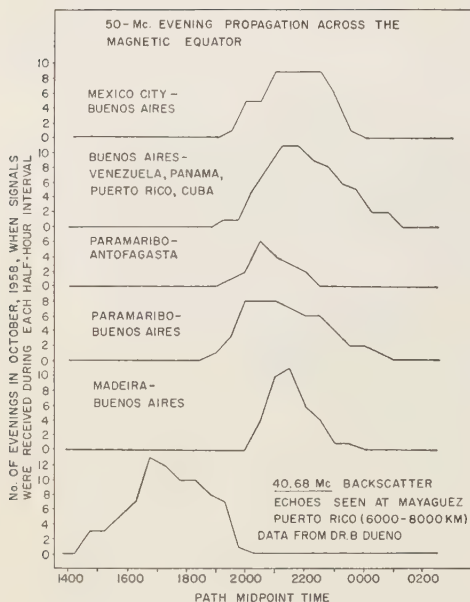


Fig. 4. Diurnal variations observed on transequatorial paths in South America.

nals (in this case a 24-hour trolley repair paging station in Montevideo) as beacons. This is just one example, but it may help inspire confidence in what follows.

Diurnal variations. The top five curves of Figure 4 show the diurnal characteristics observed during October 1958 on specific 50-Mc/s TE paths in South America. The bottom curve shows the same thing for the 40-Mc/s long-range echoes seen by Dueño at Puerto Rico. With time divided into half-hour intervals, a count was made of the number of evenings signals were received on each path during each interval. Then during plotting the curves were adjusted to the same mid-point time scale. The 50-Mc/s curves show similar 2100–2200-hour maxima and agree well with the 2130 mid-point time peak reported by Cracknell [1959] for 50-Mc/s TE between Southern Rhodesia and Cyprus. Results on 40 Mc/s are markedly different. Note that the 40-Mc/s transequatorial echoes faded out at Puerto Rico before Buenos Aires amateurs began to reach Puerto Rico on 50 Mc/s. If the same mechanism accounts for both, it must be a very unusual one, especially

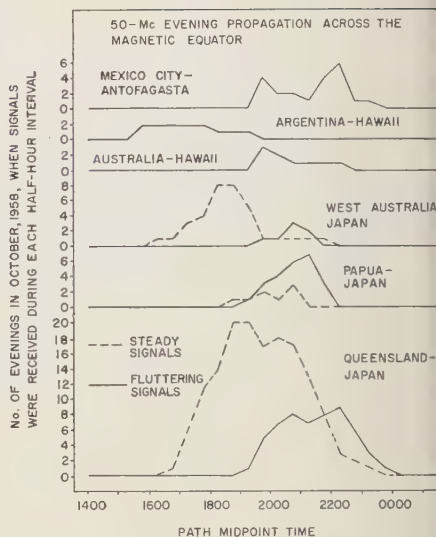


Fig. 5. Diurnal variations of transequatorial paths around the world.

since the 40-Mc/s echoes do not appear again later in the evening.

In Figure 5, additional 50-Mc/s transequatorial paths have been handled identically to those of Figure 4. At the top is one more South American path for comparison purposes. This is the Argentina to Hawaii. This is the only 50-Mc/s circuit whose diurnal variation distinctly resembles that of the 40-Mc/s experiment. They may share causative mechanisms as well. The third curve, Australia to Hawaii, appears to be standard 50-Mc/s TE. The bottom two curves, all Japan to various parts of Australia, have two parts. Whereas flutter fading is always present on South American TE, Japanese and Australians have steady signals as well. The dashed curves show the steady signals, which tend to appear first and may resemble the Puerto Rico backscatter. Solid curves are for the fluttering signals, which seem similar in behavior to the other TE paths.

Diurnal variations of several medium-range paths are shown in Figure 6. All but the bottom curve are from 50-Mc/s amateur results. The y-coordinates are the same as before. Five of the circuits are across South America; the sixth is Japan to Manila. All show about the same

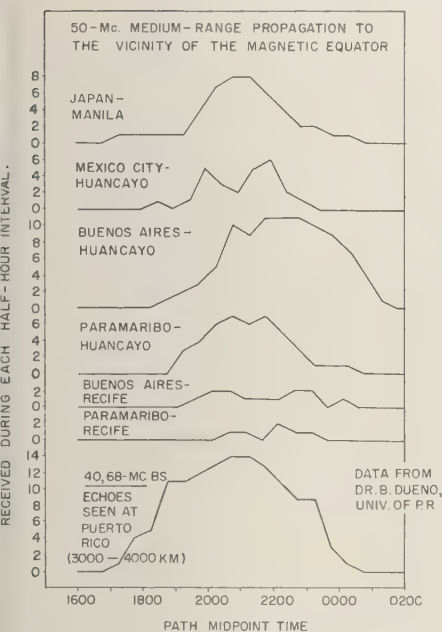


Fig. 6. Diurnal variations of several medium-range paths.

00-2200-hour maximum that was seen for transequatorial work. Interestingly enough, so does the curve drawn from Dueno's 40-Mc/s medium-range backscatter results.

The last diurnal plot, one for short-range propagation, appears as Figure 7. Again, path midpoint time is plotted along the abscissa, and the top two curves are number of occurrences per month as a function of this time. The upper curve is for Buenos Aires to Antofagasta (using NBS station CESAE at Antofagasta as a beacon) during October 1958. The middle curve is for the same month in 1957, but CESAE was not on the air at that time so only results from Argentina, Chile, and Uruguay have been lumped together to get a reasonable amount of data. A plot for October 1957 was wanted particularly for comparison with the lower graph, taken from one appearing in the *Journal of Geophysical Research* for April 1959. This is for the anomalous Okinawa-Philippines mode, and the ordinate scale is in terms of signal strength rather than frequency of occurrence. Nevertheless, there is a strong similarity,

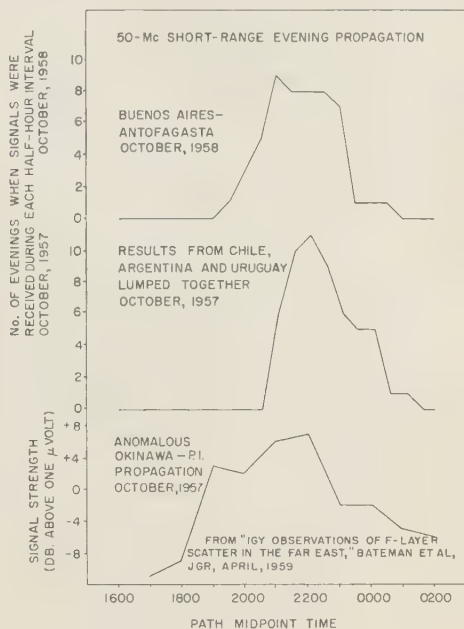


Fig. 7. Diurnal variations of short-range paths.

and the optimum and closing times are practically identical to TE and medium-range propagation.

The relation between seeming 'sporadic E' and longer modes reaching out to some 9000 km may seem unlikely. It should be emphasized that the paths discussed were chosen because substantially complete coverage was available at both ends during the evening hours. Combined 50-Mc/s amateur activity and continuous operation by commercial or scientific stations immediately below 50 Mc/s were at least comparable with the PZ1AE observing schedule shown earlier. The study utilized all intercepts reported between midafternoon and early morning for each path. TE, medium-, and short-range signals turned out to begin, be at their best, and end at almost the same mid-point times. All, more often than not, have similar fading characteristics. All are observed on the same evenings. Because of the times these modes are operative they seem less likely to involve layer tilts than the transequatorial echoes seen at the Virgin Islands and Puerto Rico on lower frequencies. The amateur work

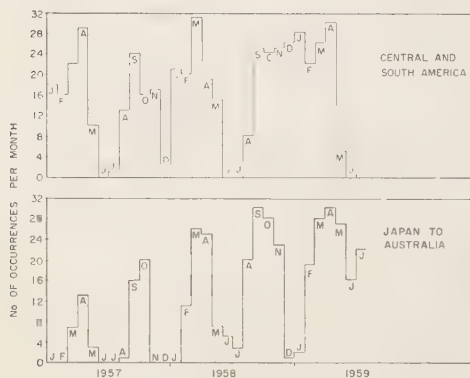


Fig. 8. Seasonal variations of evening transequatorial propagation in the Americas and the Far East.

appears more closely connected in time with the patch of high electron density recently observed [Gates, 1959] over Bogotá (magnetic dip angle 32°) about 2200 hours. This patch, with a critical frequency around 21 Mc/s, can account for a few circuits directly. If it were known whether, when, and where such patches formed in the Far East and south of the equator in South America, perhaps they could explain some additional paths.

Seasonal variations. To observe what happens to the equatorial modes throughout the year, see Figure 8, a plot of the number of TE nights per month for two major TE areas, the Americas and the Far East, from January 1957 through July 1959. Seasonal graphs for medium- and short-range propagation look much the same. The only part of either plot that cannot be taken quite at face value is the first 9 months or so of 1957 for Japan to Australia. Until October of that year, Australian amateurs were not able to use the 50-Mc/s band. This meant that all contacts with Japan had to be cross-band between 28 and 50 Mc/s. Not too many stations were equipped for this kind of operation, and propagation conditions were not often the same on both bands. The result was fewer TE contacts than would otherwise have been made.

Past references to TE propagation have mentioned equinoctial maxima and solstitial minima. Here it appears that the situation is not quite so simple. Certainly the months of March and

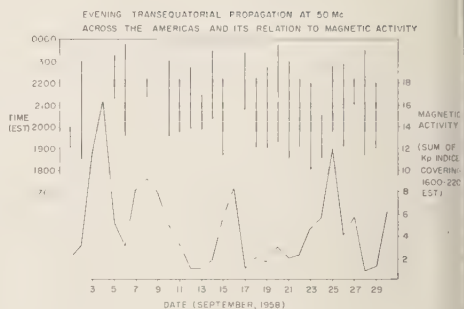


Fig. 9. South American transequatorial propagation during September 1958 as related to magnetic activity.

April and September and October are often the best, but December and January 1958 and 1959 were even better in South America, and June and July 1959 were not far down in the Far East. A critical look back over the solstices will show that November through February is almost always better than May through August in the Americas, whereas the opposite is true between Australia and Japan. One obvious thing to connect this with is the meandering of the magnetic equator. On this side of the world it is south of the geographical equator; in the Far East it is to the north. Perhaps the better solstice for transequatorial scatter is the one when it is summer along the magnetic equator. Wells [1954] reported the same thing about equatorial spread *F* during the period 1938–1954. Dueño's 40-Mc/s data for 1958 and 1959 do not reveal a solstitial preference.

Effect of magnetic activity. Figure 9 shows the interval on each evening of September 1958 when 50-Mc/s TE propagation was observed somewhere in South America. Dates run across the abscissa; Eastern Standard Time (roughly the average mid-point time for South American paths) is the left-hand ordinate scale. A diurnal curve very similar to the ones already shown can be visualized from these lines. As explained in the scale on the right, two K_p indices have been summed to obtain a figure of magnetic activity covering 1600–2200 EST for each evening. This interval includes the start of every opening and all of many. September is one of the best TE months, yet, almost every time there was high magnetic activity, TE was either

inexistent or did not begin until the optimum hour was reached. When TE conditions are less favorable a very mild disturbance will suffice to terminate the evening's 50-Mc/s contacts. Some may anticipate this next sentence: Inverse correlation with magnetic activity is a well known characteristic of equatorial spread F . Long-range evening echoes on 40-Mc/s from Puerto Rico, however, do not show this relationship.

Frequency and fading characteristics. The upper frequency limit of these equatorial modes is much above 50 Mc/s. Police signals from Cyprus on 70 to 72 Mc/s are received in the Bahamas just about as often as 50-Mc/s amateur transmissions. Television signals as high as 120 Mc/s (channel four) are frequently propagated from Caracas, Venezuela, to Argentina. The NBS group has extrapolated that their Far East anomaly might be identifiable up to 80 or 100 Mc/s. This is quite consistent with the fact that Brazilian TV channels three, four, and five (60–82 Mc/s) are often received by Argentine stations over similar paths. Any future professional investigations of these phenomena should certainly think in terms of frequencies above 50 Mc/s rather than the traditional lower ones.

Most TE signals (and very often medium- and short-range ones) have a rapid flutter fading characteristic, which seems to be of the order of 5 cycles per second. It is very deep as a rule and capable of modulating a strong signal right into the noise level. Not too much more can be said as yet, but an amateur recording program has been set up that should yield typical records suitable for more detailed analysis.

Conclusions. Transequatorial scatter propagation as observed by amateurs at 50–75 Mc/s is intimately connected with medium-range openings to the neighborhood of the magnetic equator and also with short-range work beyond a few hundred kilometers. All occur most frequently at 2100–2200 hours path mid-point time and are most pronounced around the equinoxes. All have similar fading characteristics. They are probably not related to the long-range echoes seen on lower frequencies at the Virgin Islands and Puerto Rico, where layer tilts are believed to be the cause. The amateur 50-Mc/s results have a diurnal variation similar to that of the 40-Mc/s medium-range echoes seen at Puerto Rico. As the amateur equipments

are considerably more sensitive than the backscatter sounder, however, different mechanisms may still be involved. One might be the high electron density patch that has shown up at 2200 hours in the NBS 75° meridian true height profiles. Transequatorial scatter has some of the same general characteristics (diurnal and seasonal variation and relation to magnetic activity) as equatorial spread F .

Acknowledgment. First thanks should go to the hundreds of radio amateurs who contributed so much of their time to make this program possible. Then there was the co-operation of the ARRL staff, notably Mr. Edward P. Tilton, with his uncanny ability to remember everything hams have ever done on VHF. The helpful conversations with Dr. O. G. Villard, Jr., of the Stanford Radio Propagation Laboratory, and the exchanges of data with Dr. Braulio Dueño, of the University of Puerto Rico, are acknowledged. Both collection and analysis of all data have been sponsored by the United States Air Force Cambridge Research Center under Contracts No. AF19 (604)-2171 and AF19 (604)-5235.

REFERENCES

- Bateman, R., J. W. Finney, E. K. Smith, L. H. Tveten, and J. M. Watts, IGY observations of F -layer scatter in the Far East, *J. Geophys. Research*, **64**, 403–405, 1959.
- Cracknell, R. G., Transequatorial propagation of VHF signals, *QST*, **43** (12), 11–17, 1959.
- Dueño, B., Peculiarities and seasonal variations of TE backscatter echoes observed at Mayaguez, Puerto Rico, *Research Rept. 2*, Contract AF 49 (638)-172, University of Puerto Rico, 1959.
- Ferrell, O. P., Very-high-frequency propagation in the equatorial region, Paper presented at the April 1951 URSI meeting, Washington, D. C., and abstracted in *Proc. IRE*, **39**, 719, 1951.
- Gates, D. M., Preliminary results of the National Bureau of Standards radio and ionospheric observations during the International Geophysical Year, *NBS J. Research*, **63D**, 1–14, 1959.
- Southworth, M. P., A look back and ahead at PRP, *QST*, **43** (6), 48–49, 1959.
- Tilton, E. P., The world above 50 Mc, *QST*, **31** (10), 56–57, 1947.
- Tilton, E. P., The world above 50 Mc, *QST*, **39** (12), 99–101, 1955.
- Villard, O. G., Jr., S. Stein, and K. C. Yeh, Studies of transequatorial ionospheric propagation by the scatter-sounding method, *J. Geophys. Research*, **62**, 399–412, 1957.
- Wells, H. W., F -scatter at Huancayo, Peru, and relation to radio star scintillations, *J. Geophys. Research*, **59**, 273–277, 1954.

(Manuscript received November 12, 1959; revised December 6, 1959.)

Short-Wave Fadeouts without Reported Flares

HOWARD DEMASTUS AND MARION WOOD

*Sacramento Peak Observatory
National Bureau of Standards
Boulder, Colorado*

Abstract. Short-wave fadeouts were sought which had no associated $H\alpha$ flare. To carry out this study, Sacramento Peak flare patrol films were re-examined for those days on which short-wave fadeouts (SWF's) had occurred during patrol times without reported flares. Fifteen such fadeouts were reported during the IGY period. Upon re-examination of the films, 12 of the SWF's were found to be in close time association with outstanding $H\alpha$ events, usually in the nature of pronounced plage brightenings. In 2 other cases, seeing was so poor that no definite statement could be made about possible $H\alpha$ activity. In only 1 case was there definitely no unusual $H\alpha$ event in time association with the fadeout.

For each of the 15 cases, the flare film for the entire day was re-examined without previous knowledge of the time of the SWF. Any periods of outstanding $H\alpha$ activity were selected and their times noted; only after their independent selection were their times matched to SWF times. Usually there was only 1 such outstanding period of optical activity per day. The time association with SWF was, in general, very close.

The results indicate that outstanding $H\alpha$ activity is essential in time relation with SWF occurrence.

It is known that the majority of solar flares do not produce fadeouts. The inverse relation, however, has been believed exact: that every fadeout is associated with an $H\alpha$ flare. Often such a flare has not been observed, but lack of flare patrol coverage or poor seeing has made it impossible to be sure that no flare actually took place.

Baraboshev and Gordon [1954, 1958] have referred to a number of short-wave fadeouts for which no flare had been reported, although the periods in question had been covered by flare patrol observations. These authors postulated the occurrence of what they called 'invisible flares,' that is, SWF-producing radiation not accompanied by excess radiation in $H\alpha$. The periods they covered were from January 1949 to June 1950 and the first 9 months of 1956.

During the IGY, flare patrol coverage became much more complete than for any previous period. Also, fadeout categories had been broadened to include 'gradual' and 'slow' fades, as well as the classical 'sudden' events. It was believed that with such a wealth of flare and fadeout data it might be possible to establish firmly whether or not fadeouts occasionally do occur without unusual $H\alpha$ activity. To carry

out this study, SWF's were sought for which no associated flare had been reported and which occurred when the Sacramento Peak flare patrol was operating. It was required that the fades be extensive in effect; that is, that they have a widespread index of at least 3 as re-

TABLE 1. SWF's without Reported Flare

Date	Beginning UT	Ending UT
1957 July 05	2202	2228
Aug. 11	1716	1738
Sept. 07	1717	1740
Sept. 16	1820	1835
Sept. 20	1807	1818
Sept. 21	1930	1950
Sept. 29	1400	1455
Nov. 12	1835	1857
1958 Jan. 03	1934	2020
Feb. 02*	1707	1745
Feb. 15	1628	1740
Apr. 24	1900	1950
June 02	1950	2015
July 04†	1730	2000
July 04†	2017	2138

* No outstanding $H\alpha$ event.

† Poor seeing.

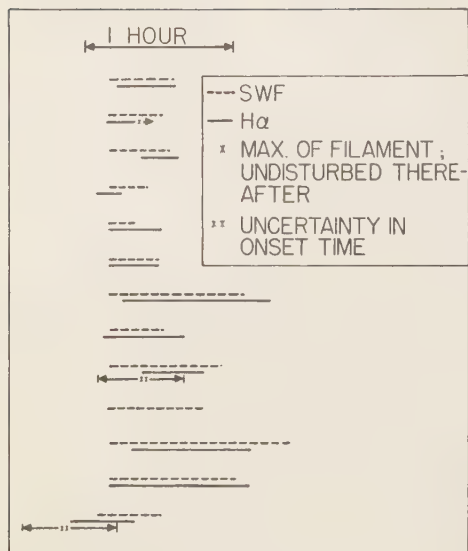


Fig. 1. Time relations between SWF's and $H\alpha$ events.

ported in the CRPL F-Series, part B. The IGY period was covered, during which about 280 fadeouts were associated with Sacramento Peak flares. Only 15 SWF's occurring during patrol times were not accompanied by a reported flare; they are listed in Table 1.

The flare films for the days on which these 15 SWF's occurred were then re-examined in detail for $H\alpha$ activity of a nonflare type. In order to avoid bias, the times of the SWF's were not noted; the whole day's flare film was studied independently of the fade time. Any period during the day that showed distinctive $H\alpha$ activity was carefully described and its duration noted. Only after such a period was selected for its $H\alpha$ importance was its time compared with SWF time.

The results show striking time correspondences between the fadeouts and these independently selected periods of optical activity. For only 1 of the 15 SWF's was there definitely no outstanding $H\alpha$ event in time association. For 2 others, seeing was so poor that no definite statement could be made about possible $H\alpha$ activity. In the remaining 12, outstanding $H\alpha$ activity occurred in close time correspondence with the reported fade times. In 10 of

TABLE 2. Fadeouts with Associated $H\alpha$ Activity

Number of Fadeouts	Type of $H\alpha$ Activity
~280	Reported flare
9	Pronounced plage brightening
1	High-speed bright feature
1	Outstanding surge
1	Absorption event
2	Poor seeing
1	No unusual activity

them only one period had been selected during the day as being of outstanding $H\alpha$ importance; thus the association with SWF time is especially striking. In the other 2 among the 12, two events of importance were selected during the day, one of which was SWF-associated. The two events in each case, however, were separated by several hours.

SWF- $H\alpha$ time correspondences are shown in Figure 1. The pairs of horizontal lines represent SWF durations and intervals of outstanding $H\alpha$ activity. For simplicity of presentation, the beginning times of all the SWF's have been lined up vertically. The one single line represents the exceptional SWF without optical activity; the 2 cases of poor seeing are omitted.

In most cases, the $H\alpha$ activity was in the nature of an outstanding plage brightening. In one, there was an extremely active surge; in another, an outstanding high-speed bright feature. It would be of interest to describe in more detail one particularly striking event which alone among the 12 cases, was an example of outstanding $H\alpha$ absorption rather than of emission. On the day of the fadeout, a very large dark filament was on the disk, of about 0.5 solar radius in length. It began a slow disintegration and completed a total disappearance 24 minutes before SWF time. Suddenly, 1 minute before fadeout beginning time, the filament began a rapid re-formation, reaching maximum in 15 minutes, 8 minutes before SWF ending time. The filament remained essentially undisturbed for the rest of the day.

Table 2 lists the total number of fadeouts occurring during Sacramento Peak flare patrol times together with the various types of associated $H\alpha$ activity.

Besides discussing the $H\alpha$ events, it would be of interest to describe a remarkable coronal event which occurred on December 20, 1958. On this date, Sacramento Peak coronal films showed an extremely bright and active green coronal region. The loop structure was very pronounced and showed large motions for coronal material. A bright yellow coronal line was also observed. Such coronal activity is typically associated with important $H\alpha$ flare activity, but in this particular case no flare occurred. In the light of $H\alpha$, the limb region in question did show large and active loop prominences, but they were well below flare intensity, indicating the relative faintness of the $H\alpha$ radiation.

According to the hypothesis that $H\alpha$ activity is not essentially associated with fadeout, but rather some other form of radiation occurring as part of the total solar event, one would expect such outstanding coronal activity to be associated with a SWF. However, no fade occurred near the time of the coronal event. Although this negative case does not provide proof, it does strengthen the hypothesis that outstanding $H\alpha$ activity is essential in time relation with SWF occurrence.

The 1 fadeout among the 15 which was definitely without important $H\alpha$ activity appears to be a real exception. On the day in question, there were two regions of moderate interest

on the disk. One, a large, complex plage region, did fluctuate and showed several surge-like events during fadeout time, though they had slower line-of-sight velocities than the average surge. In addition, a region of small filaments was mildly disturbed throughout the period of interest. However, there were other intervals during the day that were just as active. In summary, the period showed no energetic $H\alpha$ activity of any real interest.

In spite of this one exception, the results strongly indicate that virtually all short-wave fadeouts are associated with pronounced $H\alpha$ activity, onset and ending times closely corresponding for the two types of events.

Acknowledgments. We should like to thank Drs. C. S. and J. W. Warwick for their helpful suggestions regarding this study, and Dr. Helen Dodson Prince for re-examining certain of the McMath Observatory flare patrol records.

REFERENCES

- Barabashev, N. P., and I. M. Gordon, Chromospheric flares of a peculiar type and their geophysical consequences, *Dokl. Ak. Nauk. RSS Ukraine*, 1, 9 pp., 1954.
Barabashev, N. P., and I. M. Gordon, Invisible flares and proof of their reality by the rocket observations of the short-wave radiation of the sun, paper presented at the 10th General Assembly of the International Astronomical Union, 1958.

(Manuscript received October 16, 1959.)

Correlated Micropulsations at Magnetic Sudden Commencements

W. K. BERTHOLD, A. K. HARRIS, AND H. J. HOPE

*U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey*

Abstract. By the use of very large loops of wire, many square miles in area, nearly simultaneous micropulsations were recorded in Arizona and New Jersey during the initial phase of a magnetic storm. Evaluation of the two records (15 minutes' recording time) by cross correlation showed a time difference in arrival of common signals of 2 to 3 seconds. A similar difference was observed for the sudden commencement itself.

The method used is simple and relatively unaffected by interference. The maximum sensitivity, 1.7×10^{-4} gamma sec⁻¹ mm⁻¹, could be used only during magnetically quiet periods.

At two other sudden commencements, a remarkable resemblance was found in the fine details at the variations present. The over-all study includes effects from both natural and man-made causes. Another paper will discuss the man-made effects.

Introduction. One of the long-standing questions in geomagnetism is the degree of simultaneity of world-wide sudden commencements. It has been under study for 80 years, and although interesting new results have been reported by Gerard [1959], there is still no definite world-wide picture of magnetic-storm onsets. Despite the adoption, by the International Union of Terrestrial Magnetism and Electricity meeting at Madrid in 1924, of an agenda item calling for 'observing . . . times of occurrence of sudden commencements with . . . high precision in the determination of time,' geophysicists have not yet determined the world-wide time relationships with sufficient accuracy to formulate a generally acceptable theory.

A second question that has concerned geophysicists for years is the geographical extent of coherent magnetic fluctuations. On this subject there is considerable conflict in the literature. Although early results, such as those of von Engelhardt [1913], had indicated world-wide occurrence of micropulsations, later observations showed this conclusion to be too broad. More recently, for example, Duffus and Shand [1958] reported little correlation, in general, between micropulsations recorded on opposite coasts of Canada, with, however, the occasional simultaneous occurrence of similar signals at the two locations.

Current interest in these questions has increased as the result of the following develop-

ments: the postulation by Alfvén [1942] of the existence of magnetohydrodynamic (hydro-magnetic) waves in the solar atmosphere, the possibility of such waves propagating in the ionosphere [Dungey, 1954a], the use of rockets and satellites for making direct measurements at high altitudes, the observation of effects produced in the ionosphere by ground-level nuclear detonations [Daniels and Harris, 1958], the geophysical effects of high-altitude nuclear detonations [Kellogg, Ney, and Winckler, 1959], and the study of whistlers by Storey [1953] and many others. These developments may cast new light on the causes of geomagnetic fluctuations and the manner of initiation of magnetic storms.

We started our investigations (1) to study micropulsations and other fluctuations with periods of the order of seconds, (2) to ascertain whether there is coherence between fluctuations recorded at widely separated locations and, if there is, to determine the time relationship, and (3) to detect geomagnetic-field perturbations that might result from man-made causes, such as rockets or large explosions. Analysis of some of the data from our preliminary experimentation has led to the interesting results presented herein concerning micropulsations at sudden commencements and shortly thereafter. Material on possible man-made perturbations will be the subject of a separate paper.

Design of experiment. Magnetic fluctuations

having periods of the order of seconds are best recorded by amplifying the voltage induced into a large coil, as the induced voltage is proportional to the rate of change of the axial component of magnetic intensity. The recorded amplitude is also proportional to the effective area of the coil (turns times area) and the sensitivity of the amplifier-recorder system. The limitation to the over-all sensitivity is the natural background of fluctuations; hence it is useless to employ too great a sensitivity. According to *Chapman and Bartels* [1940, p. 58], Giese in 1882 laid out a horizontal cable 12 km long in the auroral zone, where the background is every high, presumably with the looped cable connected to a galvanometer, and found the system too sensitive. Electronic amplifiers are now available for signals in the microvolt range, permitting the use of a relatively small loop (effective area of the order of 0.1 km²). Alternatively, a large loop (100 km²) could be used with a comparatively uncomplicated amplifier (for millivolt-range signals) to provide the same over-all sensitivity. These two methods entail somewhat different problems in minimizing interference and providing reliable results. For the most part, experimenters have used the former method: *Ebert* [1907] used 15 turns of 2000 square meters and a galvanometer; *Chernosky, Maple, and Coon* [1954] used a 4-turn, 300-foot-diameter, buried, shielded, horizontal loop and a sensitive d-c amplifier; *Duffus and Shand* [1958] used three mutually perpendicular, 20-foot-diameter, 1100-turn coils and d-c chopper-stabilized amplifiers; and *Campbell* [1959a, b] used 2-meter-diameter coils of 21,586 turns, also in three orientations.

Experimentation was begun using the second

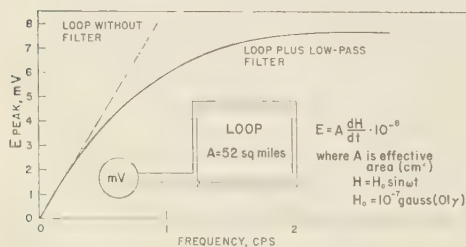


Fig. 1. Loop voltages produced by magnetic-field pulsations.

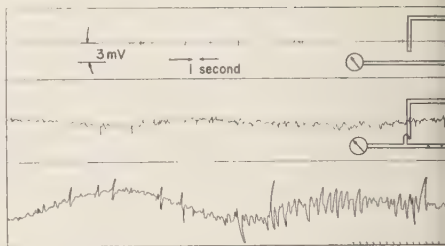


Fig. 2. Electrical-interference test and typical recording at Grand Canyon, Arizona.

method (very large loop), which has certain definite advantages. First, the large induced voltage renders insignificant such interference sources of interference as chemical or thermoelectric emfs. Second, inductance and resistance may be kept to a minimum by using but a few turns and a very large diameter. Third, a loop of large extent minimizes the effects of local interference sources, such as man-made disturbances or movements of the wire by the wind or vibrations, and it might also average out some of the local variations in the natural background. On the other hand, a disadvantage of using a very large loop is restriction to measurement of only the vertical component, which is relatively sensitive to the influence of local ground or sea conductivity (*Duffus and others*, 1959; *Sen'ko*, 1958).

Installations. Our first experimental site was in the Kaibab National Forest, just south of Grand Canyon National Park, Arizona. The terrain at this site is flat, 6000 feet in elevation, partly wooded, thinly populated, and accessible by jeep. Power lines are more than 10 miles away, and industrial activity in the nearest town is negligible. We installed 30 miles of two-conductor telephone wire (WD-1) on towers in the ground so as to enclose an area of 26 square miles, and connected this wire into a 2-turn loop of 52 square miles effective area. A sinusoidal magnetic variation with an amplitude of 1 gamma (0.01 gauss) and a period of 10 seconds would induce 0.8 mv into the loop (Fig. 1). Since the induced voltage is directly proportional to frequency, a 1-cps sine wave would provide an 8-mv input to the amplifier. To prevent frequency fluctuations above 1 cps from becoming dominant, a low-pass filter was

ed that nearly equalized the sensitivity of system in the range from 1 to 20 cps (Fig. 1) and effectively cut off higher frequencies. In addition, this filter highly attenuated the 60-cycle interference caused by our generator. Using a rush recording system that included a chop-stabilized d-c amplifier with a gain of 1000, we obtained a maximum voltage sensitivity of 0.1 mv/mm. Thus the over-all system enabled us to detect a 1-cps sinusoidal variation with an amplitude of 10^{-9} gauss.

The two-conductor arrangement permitted us to make a simple check to determine whether the measured voltage is produced only by changes of the magnetic field or whether it is produced in part by extraneous electrical variations. A simple change of connections reduces the effective area of the loop to zero without requiring anything else (Fig. 2), and changes the magnetic field are thus removed as a source of the voltage variation. A straight line should be recorded if electrostatic-field variations, thermoelectric emfs, electrochemical emfs, etc., are negligible. The upper part of Figure 2 shows such a line. The small short-duration peaks may be due to electrical discharges. Directly below the upper part of Figure 2 is a recording made immediately thereafter with the wires reconnected to form a 2-turn loop. Here the signal caused by the changes of the magnetic field is obviously quite different from the upper signal. This test indicates, then, that electrical interference is not appreciable for a loop of such large dimensions. The third part of the figure shows a more typical recording, made on another day, which includes a long-period (about 35 seconds) variation, some 1-second pulsations, and superimposed irregular fluctuations. Several months later, two similar, but smaller, loops were installed adjacent to each other in a sparsely populated region of New Jersey (Wharton Tract, Burlington County). The smaller loop had an effective area of 12 square miles; its center was about 5 miles south of the center of the large loop, which had an effective area of 23 square miles. These two loops gave us proof that errors introduced by equipment and by near-by interference comprise a negligible part of the record. Readings made simultaneously have essentially the same shape. Small differences in phase and

amplitude for frequencies of 1 cps and upward indicate the desirability of a further investigation of the causes, which we plan to make. At any rate, the results indicate that shielding is not required for loops of great extent.

We might mention some of the unforeseen difficulties of our experiment as well as possible ways of avoiding them in the future. First, the wire was frequently broken by animals or automobiles. More careful installation will prevent such accidents. Second, during wet weather, ground leakage and chemical action sometimes resulted in a d-c signal. Better installation will help here too, and we can overcome residual trouble by using balanced differential amplifiers. Third, lightning within a range of about 20 miles caused a large-amplitude spike lasting only a fraction of a second, but requiring reduction of the sensitivity setting. Unfortunately, thunderstorm activity is high in the areas we chose, particularly in summer. Isokeraunic maps show that site selection could be improved manyfold in this respect.

The Arizona loop, installed in May 1958, was operated daily during the 12 daytime hours until August. From the beginning of August until early September, both the Arizona and New Jersey loops were operated around the clock.

Observations and results. On magnetically quiet days, the general character of the records was markedly different at the two stations. In Arizona, pulsations that were almost perfectly sinusoidal were common during the daytime. Their periods, usually in the neighborhood of 35 seconds, ranged from 15 to 40 seconds. These pulsations continued for several hours at a time, but with a varying amplitude sometimes as great as 0.5 gamma. In New Jersey, the variations were in general irregularly shaped with quasi-periods of 10 seconds to several minutes.

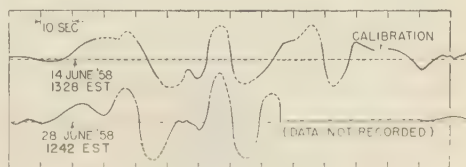


Fig. 3. Correlation of two sudden commencements recorded at Grand Canyon, Arizona.

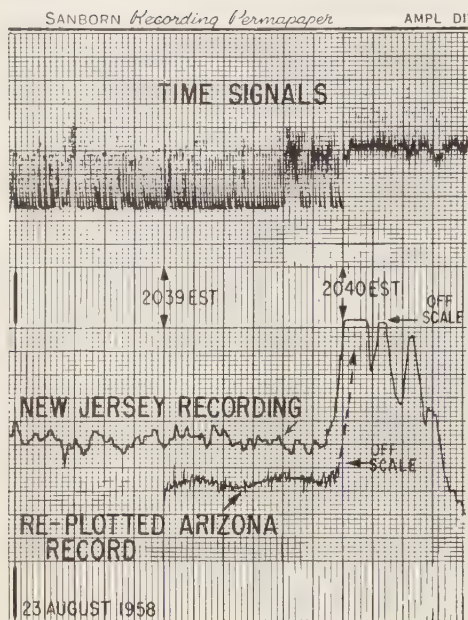


Fig. 4. Actual recording from loop at Wharton Tract, N. J., showing sudden commencement a few seconds before 2040 EST, with the recording made at Grand Canyon, Arizona, replotted on the same time scale.

During magnetically disturbed periods, however, pulsations were more regular. Both the Arizona and New Jersey observations are in good agreement with published results for data obtained by means of smaller coils and magnetic variographs.

Most magnetic storms start with a sudden commencement, but the details of the shape vary widely and are largely a function of local time. Sometimes the onset is gradual, which makes it difficult to assign a precise time to the sudden commencement. An interesting pair of such sudden commencements, occurring fortuitously only 2 weeks apart and at the same time of day (within an hour), were recorded at our Arizona installation (Fig. 3). The curves show obvious similarities. A 30-second period and a superimposed 15-second period are evident.

The sudden commencement that occurred August 24, 1958, at 0140 GMT (2040 EST August 23), had a very rapid onset, as can be

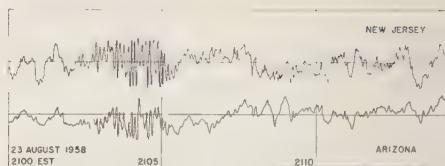


Fig. 5. Replotted simultaneous data records at two loops during initial phase of magnetic storm. (Sudden commencement occurred at 2040 EST.)

seen in Figure 4, which compares the actual recording at our New Jersey station with the data from our Arizona loop replotted to the same time scale. The relative amplitude was selected to make the later variations of about 1-minute period at the two stations approximately equal. At both stations the pen went off scale before the operator could adjust the scale factor. Nevertheless, enough can be seen to indicate that the sudden commencement occurred from 1 to 3 seconds earlier at New Jersey than at Arizona. But, as this was unsatisfactory evidence, we sought further data. Careful examination of Figure 5 shows that the time displacement continued to be in the same direction. This figure shows the striking relation between variations occurring at the two stations during the initial phase of the magnetic storm (20 to 35 minutes after the sudden commencement).

To study this indicated time delay further we decided to compute the correlation coefficient for the two curves of Figure 5, not only at time coincidence but also with various time shifts of one curve relative to the other. If a digital computer was available, we were able to use 1-second increments up to 1 minute in each direction. The resultant correlation curve for the 15-minute interval is shown in Figure 6. The outstanding feature of this curve is the 7-second fluctuation, caused by the $2\frac{1}{2}$ -minute part of Figure 5 that contains a strong, regular 7-second pulsation at both stations.¹ The correlation curve shows that this component occurred about 2 seconds later at Arizona than at New Jersey.

¹ This 7-second pulsation appears to have been a world-wide phenomenon; it was found to be present in 'rapid-run' data recorded at Chamb-la-Forêt by the Institut de Physique du Globe, Paris.

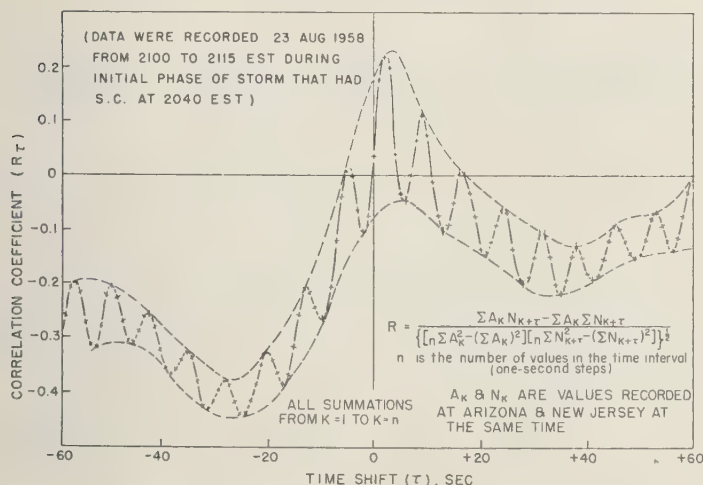


Fig. 6. Cross correlation between records shown in Figure 5.

ersey. If this component were removed from the correlation curve, the resulting curve, as indicated by the envelope, would show a delay of 3 to 3½ seconds. A periodicity of about 60 seconds is indicated by the shape of the envelope. The facts that the maximum value of positive correlation is low and that the bulk of the curve shows negative correlation indicate that there was a long-period oppositeness between the two stations during this 15-minute period.

Discussion. The few results presented above are an indication that the micropulsations occurring during the initial phase of a magnetic storm show some interesting features that justify further experimentation and more-comprehensive studies. The time comparison of a specific feature of a micropulsation at two stations is made difficult by shape differences, which usually permit accuracy no better than to the nearest second. By using the integrating operation of a cross correlation, the accuracy can be improved to the order of a tenth of a second, because the mean error decreases and the effect of the random noise becomes negligible as the interval of integration is increased. The two similarly shaped sudden commencements recorded at our Arizona station had a fundamental period of about 30 seconds. Since both of them occurred near local midday, and

the 30-second period is similar to the typical periods that were common at this site in the daytime, we are led to suppose that a resonance mechanism in the outer atmosphere [Dungey, 1954b] is involved. Hence, although the initiating source might contain fluctuations of other periods, the resonance mechanism appears to be the dominant factor in this instance.

The indications are clear that the time difference observed for the start of the sudden commencement at the two widely separated locations was still present some 25 minutes later for the micropulsations. This fact would seem to suggest that the variations are present in the source of the disturbance, perhaps having been propagated as hydromagnetic waves from the region of interaction in the outer atmosphere. The slightly longer time delay for the longer-period fluctuations is indicative of propagation differences. The negative correlation that we found between Arizona and New Jersey for long-period vertical variations (order of 15 minutes or more) is in good agreement with the findings of Matsushita [1958]. He attributes vertical-change differences between stations at sudden commencements to the induced earth current, and reports that Fredericksburg, Virginia, and Tucson, Arizona, always had opposite signs for the 15 common sudden commencements analyzed.

Acknowledgment. The results described in this paper are part of a broader geomagnetic program that was initiated by Dr. Hans A. Bomke, U. S. Army Signal Research and Development Laboratory. The objective of this program is to enhance the understanding of physical phenomena occurring in space.

REFERENCES

- Alfvén, H., Existence of electromagnetic-hydrodynamic waves, *Nature*, **150**, 405-406, 1942.
- Angenheister, G., Über die Fortpflanzungsgeschwindigkeit magnetischer Störungen und Pulsationen, *Göttinger Nachr., Math.-Physik. Kl.*, **565-581**, 1913 (see also *Terrestrial Magnetism and Atmospheric Elec.*, **25**, 26-32, 1920).
- Campbell, W. H., A study of micropulsations in the earth's magnetic field, *Science Rept. 1, Nonr 233 (47)*, Institute of Geophysics, University of California, April 1959a.
- Campbell, W. H., Studies of magnetic field micropulsations with periods of 5 to 30 seconds, *J. Geophys. Research*, **64**, 1819-1826, 1959b.
- Chapman, S., and J. Bartels, *Geomagnetism*, Oxford University Press, 1940.
- Chernosky, E. J., E. Maple, and R. M. Coon, Rapid geomagnetic fluctuations at Tucson, Arizona, *Trans. Am. Geophys. Union*, **35**, 711-721, 1954.
- Daniels, F. B., and A. K. Harris, Some effects of strong blast waves upon the ionosphere, presented to URSI-IRE Meeting at Penn. State Univ., Oct. 20-22, 1958 (abstracted in *IRE Trans.*, **AP-7**, 108, 1959).
- Duffus, H. J., and J. A. Shand, Some observations of geomagnetic micropulsations, *Can. J. Phys.*, **36**, 508-526, 1958.
- Duffus, H. J., J. A. Shand, C. S. Wright, P. W. Nasmyth, and J. A. Jacobs, Geographical variations in geomagnetic micropulsations, *J. Geophys. Research*, **64**, 581-583, 1959.
- Dungey, J. W., The propagation of Alfvén waves through the ionosphere, *Penn. State U. Ionosphere Research Lab. Sci. Rept. 57*, 19 pp., 1954.
- Dungey, J. W., Electrodynamics of the outer atmosphere, *Rept. Phys. Soc. Conf. on the Physics of the Ionosphere*, pp. 229-236, Cavendish Lab., Cambridge, 1954b.
- Ebert, H., Magnetic pulsations of short period, *Terrestrial Magnetism and Atmospheric Electricity*, **12**, 1-14, 1907.
- Gerard, V. B., The propagation of world-wide sudden commencements of magnetic storms, *J. Geophys. Research*, **64**, 593-596, 1959.
- Kellog, P. J., E. P. Ney, and J. R. Winckler, Geophysical effects associated with high-altitude explosions, *Nature*, **183**, 358-361, 1959.
- Matsushita, S., IGY pilot studies for geomagnetic data made at a closely spaced net of seven stations in the United States, *High Altitude Observatory Solar Research Mem.* **119**, November 24, 1958.
- Sen'ko, P. K., Unusual localized character of magnetic variation in the Mirny region, *Informacionnyi Bull. 1, Soviet Antarctic Expedition 'Morskoi Transport' Press, Leningrad*, pp. 81-82, 1958. (Translated by E. R. Hope, Directorate of Scientific Information Service, DRB Canada, March 1959, T319R.)
- Storey, L. R. O., An investigation of whistling atmospherics, *Phil. Trans. Roy. Soc. London, A*, **246**, 113-141, 1953.

(Manuscript received September 28, 1959;
revised November 19, 1959.)

A Proposed Experiment for the Investigation of an Energy Dependence of Photon Velocity in Vacuo

S. D. SOFTKY¹ AND R. K. SQUIRE

*Aerojet-General Nucleonics
San Ramon, California*

Abstract. It is argued that space may be a dispersive medium for electromagnetic radiation. Present experimental evidence is inconclusive. The possibility now exists for the measurement of photon velocities for energies ranging from radio frequencies to Mev gamma rays. The radiation source would be a nuclear bomb and the base line approximately 2 light seconds. The signal strengths are more than adequate.

The velocity of propagation of electromagnetic radiation is a physical constant of fundamental importance to many physical theories. The measurement of c for optical photons has taxed the ingenuity of experimental physics for half a century and, more recently, delicate and precise experiments establishing the 'wavelength times frequency' product for the microwave region have given us the most accurate values presently available. However, the velocities of lower-frequency radio waves and high-energy (X-ray and gamma ray) photons have presented such difficulties to measurement that no experiments have been performed which even approximate the precision of those optical and microwave measurements. For convenience it is assumed that these radiations have the same velocity as optical photons.

This assumption may be incorrect. If results of the optical and microwave experiments are compared a discrepancy becomes evident; either these radiations do not, indeed, propagate at the same velocity or else the experimental errors of the experiments are grossly understated. The discrepancy may amount to 16 m/sec. When it is considered that this occurs for frequencies differing by a factor of only about 10^6 , whereas the electromagnetic spectrum from radio to gamma-ray energy has a frequency spread of 10^{18} and more, then it becomes evident that there may be a really significant variation of photon velocity with energy which has not been detected. In any case, there are serious gaps in our experimental

knowledge of electromagnetic propagation which should be filled.

It is interesting to speculate about what mechanisms might yield an energy dependence of photon velocity. There may exist frequency dependent nonlinearities in electromagnetic fields, photons may have (very small) finite rest masses and move as highly relativistic particles, there may be quantum mechanical effects exhibited over certain ranges, and other reasons might occur to the reader. One of the more intriguing possibilities is that the interaction of Mev-energy photons with space and the propagation of such gammas by virtual pair production might lead to significantly different velocities for photons below, at, and above the 1 Mev pair production threshold. If some of these effects exist they have so far eluded observation, and a really precise comparison of photon velocities over a really wide energy range might prove to be the first way to study them.

Thus, it appears that over the narrow energy region where experimental measurements are available the photon velocity may not be independent of energy and that furthermore there are very wide energy regions for which there is no experimental evidence at all to substantiate an important assumption, namely, the independence of c on frequency. It is suggested that this kind of information is of such fundamental importance as to be worthy of major experimental effort.

If it is then concluded that a comparison of the c 's for different wavelengths would be worth while, it can be made plausible that such

¹ Now at Stanford Research Institute, Menlo Park, California.

TABLE 1

Type of Radiation	γ rays and X-rays	Visible Light	Radar and Radio waves
Wavelengths	Very short	Medium	Long
Reflection or refraction properties	Cannot be reflected or refracted	Can be reflected or refracted	Reflection or refraction vary over region
Min. pulse width	$w = 10^{-9}$ sec	$w = 10^{-9}$ sec	$w > 10^{-6}$ sec
Minimum base line	6 km	6 km	6000 km

a comparison is impractical in the laboratory. For a comparison between really wide ranges of wavelength one must insist upon two features of the experiment: (1) the base line must be in vacuo to avoid index of refraction corrections which are large compared with the necessary accuracy, and (2) the measurement should be made by a direct comparison of times of flight in order to avoid using features of propagation velocity which are unique to any particular range of wavelength. The resulting restrictions upon the experiment are severe if a difference as small as 16 km/sec out of 3×10^8 km/sec is to be measured with relative certainty.

If one divides the entire electromagnetic spectrum roughly into three parts as shown in Table 1, where w in the minimum pulse length of radiation attainable with present techniques, then one can assume that the minimum acceptable resolution in time of flight is w . Thus, for a base line d , and velocities c and $c + 16$ km/sec,

$$w = d[1/c - 1/(c + 16)] \simeq 16 d/c^2$$

$$d = \frac{c^2 w}{16} \simeq \frac{9 + 10^{10}}{16} w = 6 \times 10^9 w \text{ km}$$

It is clear that a base line utilizing multiple reflections to achieve the required length would be really practical only for visible light, as in the Michelson experiments, and for the very highest of the microwave frequencies. At the very best, c 's could be compared over only a narrow frequency range and the total optical path in vacuo would have to be at least 6 km. To compare c for visible light with that for X-ray and quanta would clearly require a vacuum chamber 6 km long, an attainable but difficult achievement even if adequate radiation

pulse sources could be realized. To directly compare c for either of these wavelengths with that of radio waves would require a vacuum chamber 10^8 times longer, simply because a radio wave pulse width would have to consist of many cycles, with consequent broadening of available time resolution. It is reasonably obvious that an experiment to directly and unambiguously compare c 's for the whole electromagnetic spectrum could not be done by ordinary laboratory techniques and probably not at all on earth.

It appears, however, that the experiment could be done in interplanetary space. It is a well-known fact that most of the energy of a nuclear bomb is released within microseconds and that a large fraction of this energy is released promptly as electromagnetic radiation with a broad frequency distribution. Consequently, if a large enough nuclear bomb were detonated sufficiently far from the earth, in space, the relative times of arrival at the earth of radiation of different wavelengths could be used to compare c 's for those radiations.

Actual detection of all but the optical and some of the radio frequency radiation would have to be by instruments located above the sensible atmosphere because of absorption effects. These instruments could be carried by an earth satellite. The base line would be approximately known by triggering the detonation with a radio pulse from the earth and measuring the time lapse between transmission and the mean time of reception of the bomb's radiation.

The relationship of intensity versus time of the various frequencies of radiation emitted by a bomb would have to be known with precision. This could be measured by detecting these radiations from a near-by explosion (in space,

which the short time base would not permit significant amount of velocity dispersion.

One assumes a 1 megaton fission bomb exploding 6×10^5 km from the earth, one can calculate the following signal strengths incident on a detector just above the atmosphere:

Gamma energy (photons/cm²) = 6×10^3
 assuming 2 γ 's per fission)
 Thermal energy (photons/cm²) = 2×10^7
 assuming a 0.01 per cent luminous efficiency)
 Energy (watts/cm²) = 10^{-9}
 assuming 10^{-8} of the energy released as RF per microsecond)

These signals are orders of magnitude above detection limits of ordinary instrumentation. Since they would be received over a time interval measured in microseconds, the signal-background ratio would be very large.

With the assumed base line of 2 light-sec-

onds, a radiation pulse of the order of 5- μ sec width would allow the comparison of the relative values of c 's to about 0.7 km/sec or some 20 times the accuracy required to check the discrepancy of 16 km/sec out 3×10^5 km/sec.

Thus we conclude that the experiment is capable of providing the necessary signal strength and time resolution to allow the direct comparison of the in-vacuo propagation velocity of photons of widely differing energy. It appears that the technical means of doing such an experiment are now available and that the requirements would be simple additions to already planned extraterrestrial experimentation. The results of such a comparison, whether it showed an energy dependence or independence of photon velocity, would be of great intrinsic interest to science.

(Manuscript received November 16, 1959.)

A New Method of Computing the Deacon Wind Profile Parameters

FRANK L. MARTIN

*U. S. Naval Postgraduate School
Monterey, California*

Abstract. An explicit formula for the local Deacon profile parameter β is derived. This formula involves surface roughness, local Richardson number, and adiabatic friction velocity, as well as the wind shear and elevation. The development is based on the surface layer theories of Lettau as well as on the derivative form of the Deacon profile. Local values of β fluctuated considerably with height, but weighted mean values in the vertical gave reasonably good correlation with the weighted mean value of $(1 + x)^2$ in the layer, where x is essentially the Richardson number.

Introduction. Lettau [1949, 1952] has published theories of surface layer turbulence which present extensions of the Prandtl mixing length theory. In the simpler of these models, Lettau [1949] was able to describe the vertical variation of the wind in terms of three surface layer parameters: the roughness parameter, the vertical heat flux, and the surface stress. Lettau's theory gave a vertical wind profile which had qualitatively the proper dependence upon stability, as was indicated by the sign of the curvature of the wind profile plotted against the logarithm of height. However, its actual application was rather involved.

Deacon [1949] also generalized the logarithmic wind profile of the surface layer to take account of non-neutral stability. His work was based primarily upon observational data, and his result may be stated in the form

$$\frac{\partial u}{\partial z} = \frac{u_*}{kz_0} \left(\frac{z + z_0}{z_0} \right)^{-\beta} \quad (1)$$

where $u_* = (\tau/\rho)^{1/2}$. Deacon then exhibited β empirically as a decreasing function of the quantity J , defined by

$$J = \frac{\theta_5 - \theta_{0.2}}{u_1^2} \quad (2)$$

In (1) and (2), u_* is the friction velocity; k is the von Karman constant, usually taken to be 0.4; z_0 is the roughness parameter; $\theta_{0.2}$ and $\theta_{0.5}$ are the potential temperatures at the 5-meter and 0.2-meter levels, respectively; and u_1 is the mean wind at the 1-meter level. The integrated

form of (1) is

$$u = \frac{u_*}{k(1 - \beta)} \left[\left(\frac{z + z_0}{z_0} \right)^{1-\beta} - 1 \right] \beta \neq 1 \quad (3)$$

If $u_*/k(1 - \beta)$ is assumed to remain constant between two data points z_2 and z_1 , the ratio u_2/u_1 may be written

$$\frac{u_2}{u_1} = \frac{\left[(z_2 + z_0)/z_0 \right]^{1-\beta} - 1}{\left[(z_1 + z_0)/z_0 \right]^{1-\beta} - 1} \quad (4)$$

and may be solved for β only by trial-and-error methods. However z_0 must first be determined as the $u = 0$ intercept of the neutral (logarithmic) wind profile, which occurs near sunrise or sunset on clear days.

Lettau [1956] has shown that Deacon's β may be written in the form

$$\beta = - \frac{\partial \ln (\partial u / \partial z)}{\partial \ln z} \quad (5)$$

so that β is essentially proportional to the local curvature of the wind profile. An explicit, rather than implicit, solution for β based on (5) would therefore require three data points in the vertical.

In this paper, combination of some theoretical results of Lettau's models leads to explicit formulas for u_* and β which may be solved by using only two data points in vertical.

Isotropic and nonisotropic turbulence. It will be advantageous to review some of the basic concepts of Lettau's theory [1949]. According to Lettau, the Austausch coefficient $A = -\rho l^2 w'$ may be written

$$A = \rho l w^* \quad (6)$$

where l is the usual root-mean-square mixing length and \bar{w} is called, by Lettau, the *mixing velocity*. According to the turbulence theory

$$\tau = -\overline{\rho u' w'} = A(\partial u / \partial z) \quad (7)$$

and Lettau writes

$$\tau = \rho \bar{u} \bar{w} \quad (8)$$

where \bar{u} will here be called the *virtual friction velocity*, in contrast with the usual friction velocity u_* . Combining (6), (7), and (8) leads to

$$\bar{u} = l(\partial u / \partial z) \quad (9)$$

and therefore \bar{u} is essentially proportional to the standard deviation of the velocity fluctuations in the direction of the mean wind. Lettau then postulates that for neutral stability (isotropic turbulence) $\bar{u}_a = \bar{w}_a$, where the subscript a indicates adiabatic lapse. In such an atmosphere, (8) shows that

$$\tau_a = \rho \bar{w}_a^2 = \rho \bar{u}_a^2 = \rho u_{*a}^2$$

so that $\bar{w}_a = \bar{u}_a = u_{*a}$, the last of these quantities being the friction velocity in an atmosphere of neutral stability. However, in a nonisotropic atmosphere none of the equalities need hold.

The mixing velocity as a function of stability. Here the treatment is again essentially after Lettau [1949]. He makes the following assumptions:

$$\bar{w} / l = \bar{w}_a / l_a \quad (10)$$

$$\bar{w}^2 / l = u_{*a}^2 / l_a - (g' / \theta_m) (\partial \theta / \partial z) \quad (11)$$

where θ_m is the mean potential temperature of any layer. Equation (11) is an acceleration equation, the first two terms having the general form $w \partial w / \partial z$, expressed in finite differences. The last term of the right side of (11) is the buoyancy contribution to the vertical acceleration. In the later theory Lettau [1952] has modified the last term by multiplying it by $N_\theta = K_{\text{heat}} / K_{\text{mom}}$, where the K 's are the corresponding eddy diffusivities for heat and momentum. In this paper we assume $N_\theta = 1$, although if the first and second terms of (11) were known, one could determine N_θ .

Combining (10) and (11), with $N_\theta = 1$, leads to

$$l = l_a / (1 + x) \quad (12)$$

$$x = g \frac{l_a^2}{\theta_m \bar{w}_a^2} \frac{\partial \theta}{\partial z} = g \frac{k^2 (z + z_0)^2}{\theta_m \bar{w}_a^2} \frac{\partial \theta}{\partial z} \quad (13)$$

In (11) and (13), θ_m may be closely approximated by T_m , the mean temperature of the layer. should also be noted that x may be written terms of the Richardson number

$$Ri = \frac{g \partial \theta / \partial z}{T_m (\partial u / \partial z)^2}$$

as

$$x = \left(\frac{\partial u / \partial z}{\partial u_a / \partial z} \right)^2 Ri$$

Combination of (10) and (12) leads to

$$\bar{w} = \bar{w}_a / (1 + x) \quad (14)$$

Since both z_0 and $\bar{w}_a = (\tau_a / \rho)^{1/2}$ may be readily determined if a neutral surface layer wind profile is available, and since $l_a = k(z + z_0)$, it follows that \bar{w} and l are known in terms of surface layer parameters, including the parameter x .

The friction velocity as a function of stability. From (1) and (8) it follows that

$$u_* = (\bar{u} \bar{w})^{1/2} \quad (15)$$

Substitution into (15) from (9) and (10) gives

$$u_* = \left[\frac{l^2}{l_a} \frac{\bar{w}_a}{\bar{w}} \frac{\partial u}{\partial z} \right]^{1/2} \quad (16)$$

Equation (1) may be considered to be valid within sublayers of approximately the lower 20 meters for some best-fitting choice of z_0 . Combination of (16) and (1) to solve for $\partial u / \partial z$ leads to

$$\frac{\partial u}{\partial z} = \frac{\bar{w}_a}{k z_0 (1 + x)^2} \left(\frac{z + z_0}{z_0} \right)^{1-2\beta} \quad (17)$$

Comparison of (1) and (17) shows that

$$u_* = \frac{\bar{w}_a}{(1 + x)^2} \left(\frac{z + z_0}{z_0} \right)^{1-\beta} \quad (18)$$

Note that (18) includes the case of neutral stability, $x = 0$. In this case $\beta = 1$, and u_{*a} is independent of elevation in the surface layer. For non-neutral stability, u_* is a function of stability through the factor $(1 + x)^2$ and the exponent β .

Computations of β and u_ .* Johnston [1959] has

puted values of β and u_* in the layers 3 to 12, 12 to 41, and 41 to 91 feet. The data used were from the University of Texas micrometeorological installation at Manor, Texas [Hardt and others, 1950]. Wind speeds were measured with Bendix-Friez aerovanes at the 12-, 41-, and 91-foot levels. The wind data at each level are represented as an average of five readings, each of 5-sec duration, over a 1-min period. In the recorded data, wind speeds were rounded off to the nearest mile per hour and therefore were not as accurate as might have been desired.

Temperatures at the micrometeorological tower were measured with thermistors whose accuracy was estimated to be $\pm 0.10^\circ\text{C}$. There were numerous levels for which time-mean-temperature readings were available, so that mean-temperature data could be accurately interpolated to wind-data levels. The finite difference form

$$= \frac{gk^2[(z_1 + z_2)/2 + z_0]^2}{w_a^{*2}(z_2 - z_1)} \frac{\theta_2 - \theta_1}{T_m} \quad (19)$$

was used, where the subscripts 1 and 2 refer to the bottom and top of any layer.

Equation (17) may be solved for β in the form

$$\frac{1}{2} \left\{ 1 + \frac{\ln kz_0}{\ln z_0/(z + z_0)} + \frac{\ln \left[(1 + x)^2 \frac{\partial u}{\partial z} / w_a^{*2} \right]}{\ln z_0/(z + z_0)} \right\} \quad (20)$$

the mean value of β for each layer may then be computed from (20), using the finite difference value of x and of $\partial u / \partial z$ for the layer, and the layer-center height $(z_1 + z_2)/2$. Apart from the effect of the neutral wind profile within ± 6 hours of computing w_a^* and z_0 , two winds in the layer would suffice for computing β in each layer.

Values of β fluctuated randomly by as much as 0.1 to 0.2 in successive half-hours and successive layers. Johnston [1959], however, concluded that most of this variation was due to rounding off of wind speed values in the recorded data. This, of course, affects the accuracy of $\partial u / \partial z$ in (20). In order to reduce the effect of random errors in time, all values of β reported in this paper have been time-smoothed by the

operation

$$0.25\beta_{i-1} + 0.5\beta_i + 0.25\beta_{i+1}$$

where $i - 1$, i , and $i + 1$ refer to successive computations at the same level.

Equation (20) suggests that a correlation exists between the local values of β and $(1 + x)^2$, but Johnston found that the local values of β computed from (20) were still subject to rather large random variations. In order to investigate the existence of a significant correlation between β and $(1 + x)^2$ with the data at hand, it was decided to apply the correlation between the weighted mean values $\bar{\beta}$ and $(1 + x)^2$ for the entire layer from 3 to 91 feet. The weighting factors employed were proportional to the thicknesses of the sublayers (3 to 12 feet, 12 to 41 feet, and 41 to 91 feet). This vertical averaging has the effect of filtering out any height variation of β , a variation which has been discussed by Davidson and Barad [1956]. The resulting linear correlation coefficient between $\bar{\beta}$ and $(1 + x)^2$ was -0.54 , a considerable improvement over the correlation before vertical averaging.

From the discussion in the preceding paragraph, it should not be inferred that the height variation of β is inconsequential; the vertical averaging was simply employed in order to remove additional random errors, introduced by initially oversmoothed wind speed data. Presumably the correlation of local values of β and $(1 + x)^2$ could be determined more adequately from the more accurate micrometeorological data collected in the Great Plains Turbulence Project [Lettau and Davidson, 1957]. On the other hand, the parameter $\bar{\beta}$ discussed in this paper is to be regarded as a Deacon profile bulk-parameter for the entire layer from 3 to 91 feet. Lake [1952] employed a somewhat different aspect of the Lettau [1949] theory, together with wind profile data of Pasquill [1949], in order to obtain bulk values of β ; however the over-all layer thickness in Pasquill's data was considerably smaller than that considered in this study.

In Figure 1, values of $\bar{\beta}$ and $(1 + x)^2$ have been plotted as functions of time for the period January 24-25, 1950. In Figure 2, a scatter diagram of $\bar{\beta}$ versus $(1 + x)^2$ has been drawn from the data appearing in Figure 1. The curve shown in Figure 2 was drawn by eye and is in

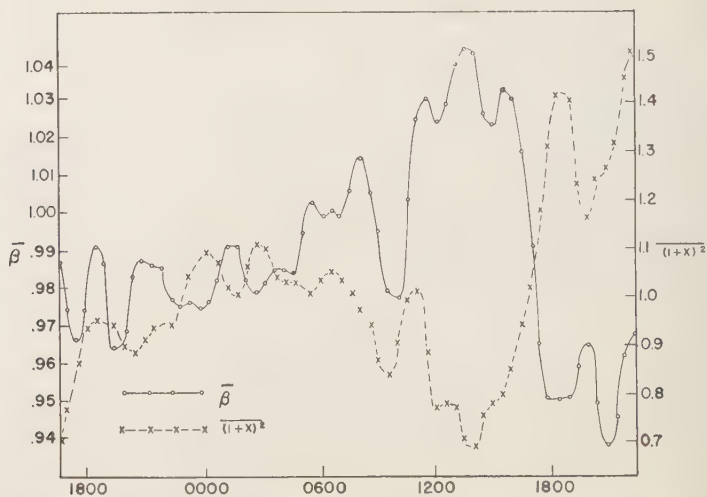


Fig. 1. Graphs of $\bar{\beta}$ and $(1+x)^2$ against time (CST), January 24-25, 1950 (after Johnston [1959]).

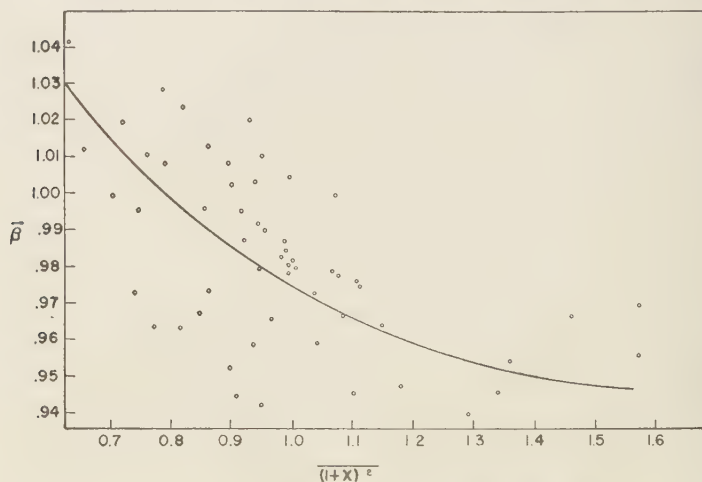


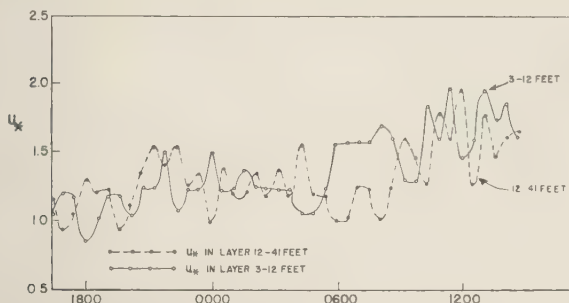
Fig. 2. Scatter diagram of $\bar{\beta}$ with $(1+x)^2$, January 24-25, 1950 (after Johnston [1959]).

many respects similar to Deacon's [1949] graph of β versus J .

The value of u_* was computed using (18). In Figure 3, graphs of u_* versus time for the period December 19-20, 1949, are shown for the layers from 3 to 12 feet and from 12 to 41 feet. There appears to be no significant tendency for u_* to vary with height from the first to the second layer. However u_* in the layer from 41 to 91 feet was systematically larger than in the

other two layers, suggesting perhaps that surface layer did not extend to the midpoint of the third layer.

If, as indicated in Figure 3, u_* is approximately constant with height in the surface layer, $(1+x)$ must be proportional to $[(z+z_0)/z_0]^{1-\beta}$. From (13), this result is qualitatively reasonable, since the factor $(z+z_0)^2 \partial\theta/\partial z$ of x is an increasing function of height in stable ($\partial\theta/\partial z > 0$) cases, whereas this factor is a decreasing function



3. Graphs of u in mph against time (CST), for the layers from 3 to 12 and from 12 to 41 feet, December 19-20, 1949 (after Johnston [1959]).

ht in unstable cases. To verify these last statements, one need only examine the con number for the vertical profile of potential temperature [cf., for example, Fig. 7.5.3 Lettau and Davidson, 1957].

Some comparisons with other research. This note gives an explicit formula, equation (20), for the local value of the Deacon wind profile parameter in terms of certain micrometeorological parameters, including the Richardson number. Lake [1952] has derived a theoretical expression for β in any sublayer 1-2 of the surface layer. His formula is

$$\beta = 1 - \left\{ \frac{\ln Y_2/Y_1}{\ln [(z_2 + z_0)/(z_1 + z_0)]} \right\}$$

where $Y = (1 + x)^2$. He also considers the possibility of a zero-level displacement. However, he computes Y by a somewhat less convenient computational formula than that just given. Both his paper and the present article indicate that β is a monotonically decreasing function of Ri , although the vertical scales of the layers considered are somewhat different. Lake's curve of β versus Ri appears to have concave-downward curvature, whereas the present paper indicates that the curve of β versus $Y = (1 + x)^2$ has concave-upward curvature. These results are not necessarily inconsistent, as one may observe by writing

$$\frac{d^2\beta}{dY^2} = \left[\beta'' - \frac{\beta'}{(1+x)} \right] \frac{1}{4(1+x)^2}$$

where $\beta'' = d^2\beta/dx^2$, $\beta' = d\beta/dx$.

Conclusions and summary. In this paper the relation between the weighted mean values $\bar{\beta}$

and $(1+x)^2$ in the layer from 3 to 91 feet has been examined. However, the nature of the variation of the local value of β with height and as a function of the other micrometeorological variables is in itself an interesting problem which deserves further investigation. The approach presented here may prove helpful in this regard.

Conversely, if local values of β are computed directly from the wind profile by either (4) or (5), equation (17) may be used to determine x . Again, on the basis of Lettau's [1952] turbulence model, the value of x may be applied to determine N_θ , the ratio of the eddy diffusivities for heat and momentum.

REFERENCES

- Davidson, B., and M. L. Barad, Some comments on the Deacon wind profile, *Trans. Am. Geophys. Union*, **37**, 168-176, 1956.
- Deacon, E. L., Vertical diffusion in the lowest layers of the atmosphere, *Quart. J. Roy. Meteorol. Soc.*, **75**, 89-103, 1949.
- Gerhardt, J. R., K. H. Jehn, and R. C. Staley, Micrometeorological research data, vol. 3, *Rept. 43, Elect. Eng. Research Lab., Texas Univ.*, 1950.
- Johnston, J. I., Use of micrometeorological data in the computation of the friction velocity, U. S. Naval Postgraduate School Master's thesis, 44 pp, 1959.
- Lake, H., A comparison of the power law and a generalized logarithmic formula in micrometeorology, *Trans. Am. Geophys. Union*, **33**, 661-668, 1952.
- Lettau, H., Isotropic and non-isotropic turbulence in the atmospheric surface layer, *Geophys. Research Paper 1*, Geophys. Research Directorate, Air Force Cambridge Research Center, Bedford, Mass., 1-86, 1949.
- Lettau, H., The present position of selected turbu-

- lence problems in the atmospheric boundary layer, *Geophys. Research Paper 19*, Geophys. Research Directorate, Air Force Cambridge Research Center, Bedford, Mass., 49-95, 1952.
- Lettau, H., Note on the structure of the atmospheric surface layer, *J. Meteorol.*, *13*, 507-509, 1956.
- Lettau, H. and B. Davidson, *Exploring the Atmosphere's First Mile*, vol. 1, Pergamon Press, New York, 1957.
- Pasquill, F., Eddy diffusion of water vapor and heat near the ground, *Proc. Roy. Soc., London, Ser. A* *198*, 116-140, 1949.
- (Manuscript received August 15, 1959; revised November 17, 1959.)

Gravimetric Determination of Ocean Tide, Weddell and Ross Seas, Antarctica

EDWARD THIEL,¹ A. P. CRARY,²

RICHARD A. HAUBRICH,¹ AND JOHN C. BEHRENDT¹

¹Department of Geology, University of Wisconsin, Madison 6, Wisconsin
and

²U. S. Antarctic Research Program, National Science Foundation, Washington 25, D. C.

Abstract. The use of the gravity meter for measurement of ocean tides is illustrated by studies on the floating ice shelves of Antarctica. The observations are complicated by high-frequency oscillations of the ice, attributed to oceanographic influences. Factors involved in the reduction of the gravimetric data are analyzed. Amplitude and phase are computed for the more significant tidal components, and the energy spectra from 0.03 to 4 cycles per day are presented. The Weddell Sea tide has both diurnal and semidiurnal components. The Ross Sea tide is diurnal, with the solar component predominating. The tidal range is greater in the Weddell Sea than in the Ross Sea. Correlation of tidal currents with changes in surface elevation provides an estimate of the inward dimension of the Ross Ice Shelf.

Introduction. Although the rise and fall of ocean waters along coasts was perhaps one of the first geophysical phenomena to be studied by early man, the measurement of tides in the open ocean remains a difficult problem for the twentieth century geophysicist. Mechanical measurements using anchor and line from surface vessels are entirely inadequate, and the use of pressure recorders resting on the bottom in deep water poses serious difficulties. It seems likely that some new technique may ultimately prove best for measurement of tides at sea.

The present paper presents a gravimetric determination of the ocean tide. The studies were made during midwinter (1957) at Ellsworth and Little America V. These two Antarctic stations were established on floating ice shelves on opposite sides of the Antarctic continent as a part of the International Geophysical Year program. Ellsworth, at 77°42.6'S, 108.0°W, on the Filchner Ice Shelf, is 2 km from the ice front and 115 km from the nearest land. The station is at an elevation of 42 meters on an ice shelf 232 meters thick. The ocean depth at the station, determined by seismic methods, is 792 meters. Little America V, at 61.9°S, 162°16.0'W, on the Ross Ice Shelf, is 5 km from the ice front but only 2 km from the embayment in the ice shelf (Kainan Bay). It is 60 km from Roosevelt Island, the nearest

land. The ocean floor at this location is 578 meters below sea level. The station elevation is 44 meters, and the shelf is 259 meters thick. Figure 1 shows the locations of the stations relative to land, and the form and extent of the ice shelves upon which they rest.

Oscillations at the ice front. The vertical movements of an ice shelf which are observed with a gravity meter are of two general types. In addition to the rise and fall of the shelf in response to ocean tides, there is a high-frequency motion of nontidal origin with periods ranging from about 15 to 50 seconds. These oscillations have been noted previously during readings of gravity meters on arctic pack ice 2 meters thick [Crory and others, 1952] and on a floating ice island 60 meters thick [Crory and Goldstein, 1957]. Figure 2 illustrates these observations with two types of gravity meters at Little America V. It is believed that the oscillations are of oceanographic origin, influenced by the ice cover. At Little America V there was extensive open water offshore in February 1958 (Antarctic summer); the tidal oscillations were so large that the pointer vibrated across the entire ocular scale and rebounded from the stops, making readings impossible. By April, with very little open water in the Ross Sea, the amplitude of oscillations had decreased to the extent that the pointer rarely struck the stops.

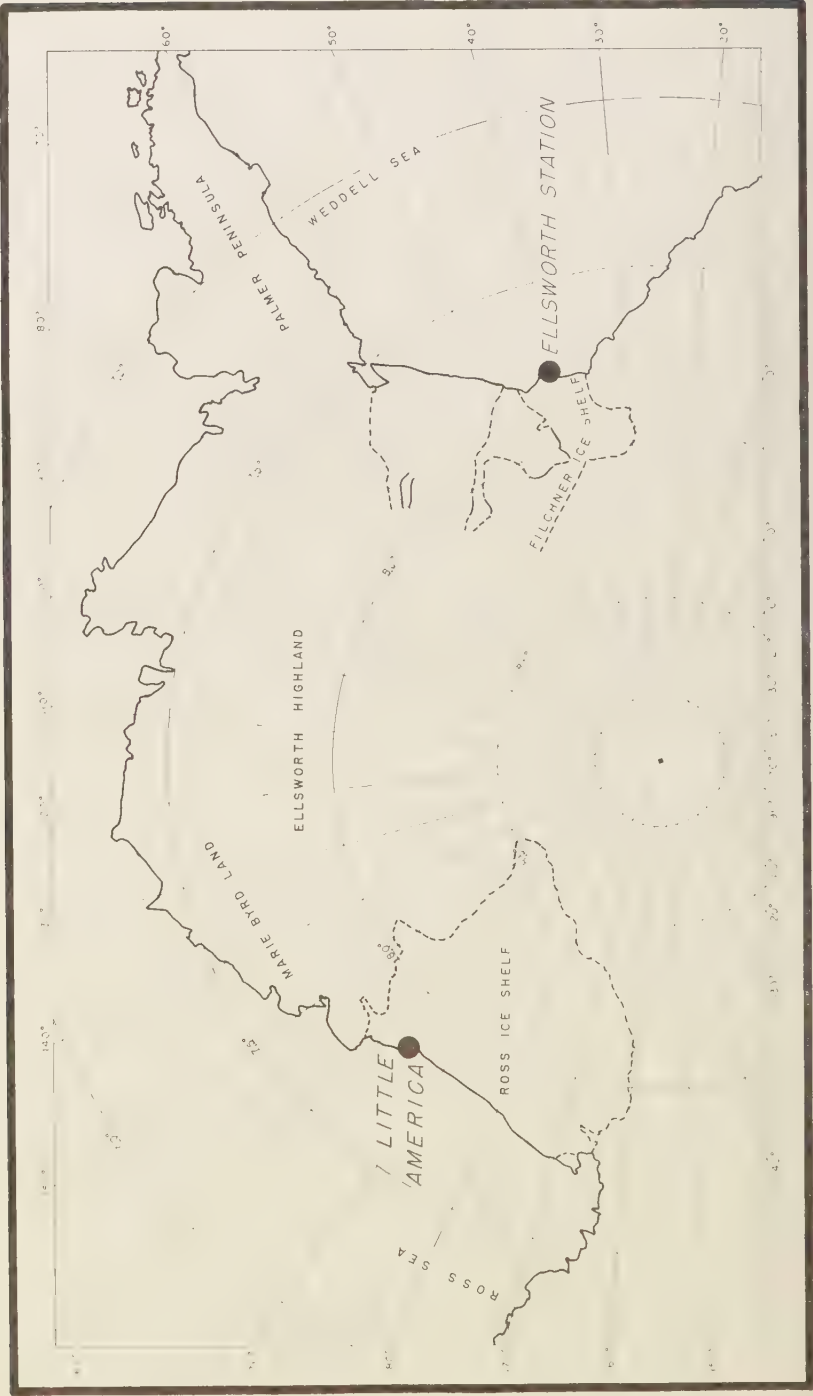


Fig. 1. Filchner Ice Shelf and Ross Ice Shelf, Antarctica.

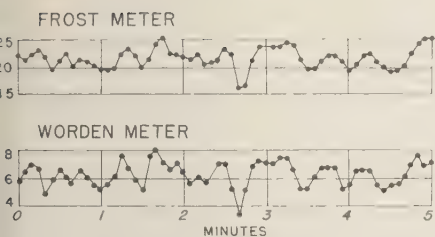


Fig. 2. Simultaneous readings of two gravity meters, March 1957, Little America V.

These high-frequency vibrations damp out very rapidly inland from the ice front. At a distance of 10 or 15 km from the edge, though still on the floating shelf, the gravity meter can be read as at a land station. Figure 3 illustrates the decrease obtained from 2 to 15 km in October 1957 when the Ellsworth oversnow traverse team proceeded inland from the station. *The tidal measurements.* The tidal observations were made during June and July 1957 at Ellsworth and Little America V with Frost gravity meters. Readings were made at approximately bihourly intervals for 1 month. The vibration of the instruments was such as to provide a reading accuracy of about ± 10 milligals, although this accuracy could not be achieved in the present instance because of the erasing process necessitated by the high-frequency oscillation of the ice shelf.

The two Frost meters had been especially selected for the oversnow traverse program because of their low drift rates. The meter at Ellsworth, for example, drifted only 0.93 mgal during the 81-day summer field season, or ± 0.34 gal/month.

In addition to the small instrumental drift, a change of gravity may be expected to result from the slow northward movement of the station during the time covered by the tidal study. The observed northward component of movement at Ellsworth is 1.25 km per year, determined by celestial navigation. According to the International Gravity Formula, this change of latitude will result in a 0.35 mgal decrease in gravity during the course of 1 month. Because of the two effects, instrumental drift and physical drift of station, are small, and because they may be expected to have opposite signs, no correction for them has been made in reducing the gravity data.

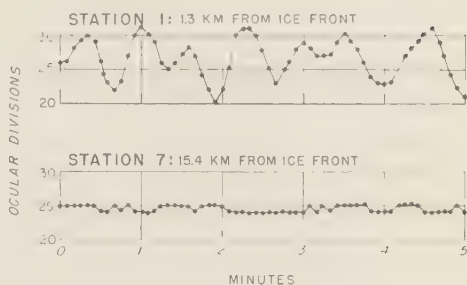


Fig. 3. Change of oscillations inland from ice front, October 1957, Filchner Ice Shelf.

The change of gravity with elevation in air at the surface of the earth is 0.3086 mgal/meter. However, in the present case, the gravity meter does not simply move up and down in air; it is also necessary to apply a Bouguer correction to allow for the changing thickness of the water column. For an infinite slab with $\gamma = 6.667 \times 10^{-8}$ cgs [Nettleton, 1940],

$$g = g_0 + h(0.0003086 - 0.00004185\rho) \text{ gal} \quad (1)$$

Since $\rho = 1.0280 \text{ g/cm}^3$ for sea water,

$$g = g_0 - 0.00026558 h \text{ gal} \quad (2)$$

Let g_0 , the mean value of gravity during the month of observation, be 0 so that the tides are measured as deviations from the mean. Then

$$h \text{ (meters)} = 3.7653g \text{ (with } g \text{ in milligals)} \quad (3)$$

The results are plotted in Figures 4 and 5.

In the gravimetric method the tide is measured with respect to the center of the earth. On the other hand, the conventional tide, as measured by a tidal gage on a coast, is the change of the sea surface with respect to the adjacent rock surface. The rock surface is subject to 'earth tides.' In principle, therefore, if the gravimetrically determined tide is to be compared with the conventional tide, a correction for earth tides is necessary.

The earth tide correction may be computed by allowing for the varying attractions of the sun and moon on the solid earth. The theoretical values thus obtained are multiplied by the factor 1.2 to allow for the fact that the earth is not perfectly rigid but yields to gravitational

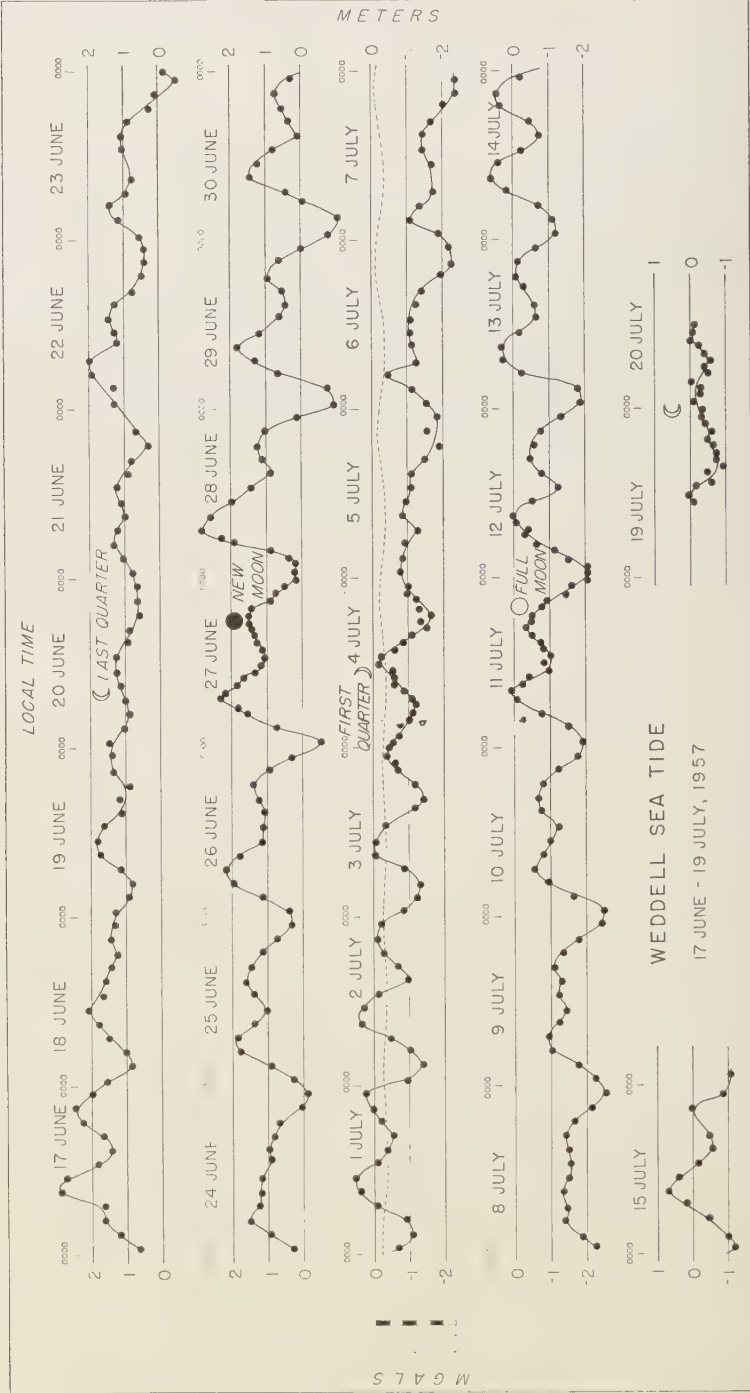


Fig. 4. Weddell Sea tide, June 17 to July 19, 1957

stresses [Baars, 1953]. Tables which give the magnitude of the earth tide are published each year by the European Association of Exploration Geophysicists as supplements to their journal, *Geophysical Prospecting*.

The Weddell Sea data have been corrected for earth tides. The magnitude of this correction is indicated by the dotted line for the third week of observation in Figure 4. Because the correction is small, the effort of computing it for the Ross Sea data was not made. The magnitude of the correction will vary with latitude, being greater at low latitudes.

It might be expected that the acceleration of the shelf itself in response to tidal movements would be reflected in the measurements, but this effect can be shown to be small. Assume a semidiurnal component involving a change in elevation of 1 meter in 6 hours. In a simply oscillating system displaced a distance x , the relationship between the maximum acceleration \ddot{x} , the period of oscillation T , and the displacement x is

$$\ddot{x} = 4\pi^2 x / T^2 \quad (4)$$

so that in the present case the maximum acceleration of the shelf is 2.12×10^{-3} mgal. From (2) the change in gravity due to elevation difference is

$$0.2656 \times 1 = 2.66 \times 10^{-1} \text{ mgal} \quad (5)$$

The two effects differ by a factor of about 10.

Data analysis. In Figures 4 and 5, the measurements are indicated by dots. These control points were connected by a smooth curve, and hourly values of tidal amplitude were read from the curve. The hourly values were smoothed by means of a numerical low-pass filter using the IBM 650 computer of the Naval Research Laboratory, University of Wisconsin. The response of the filter was such as to effectively remove energy due to frequencies above 4 cycles per day. A numerical Fourier series analysis was made of the two tidal records (3-hour smoothed values). The results after correction was made for the low-pass filter, are presented in Figure 6, which gives the energy ($\propto \frac{1}{2} \text{ amplitude}^2$) for each harmonic. The high energies at low frequencies may be partly caused by instrument drift, a low-frequency phenomenon for which no correction has been attempted. The tidal range during the period of observation are listed in Table 1.

From the hourly data, the amplitude and phase were computed for the more prominent tidal components [Schureman, 1924] on the computer. For each component, the number of hourly observations was chosen so that the total length of the record would be close to an integral multiple of the period of that component.

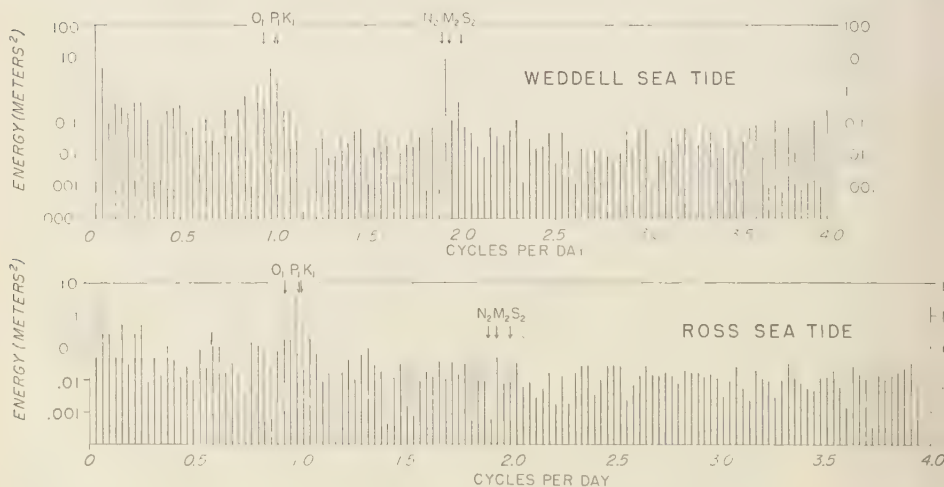


Fig. 6. Energy spectra, Weddell Sea and Ross Sea tides.

TABLE 1. Tidal Ranges

	Double Amplitude, meters	Phase of Moon
Weddell Sea tide	0.6	Last quarter
	3.2	New moon
	0.9	First quarter
	2.2	Full moon
	0.8	Last quarter
Ross Sea tide	1.4	Full moon
	0.3	Last quarter
	1.3	New moon
	0.4	First quarter
	1.2	Full moon

tidal amplitude of several harmonics on the other side of this particular harmonic were computed to provide an estimate of the noise level inherent to the data. Figure 7 illustrates this amplitude computation for the M_2 component of the Weddell Sea tide. The M_2 component is represented by the 55th harmonic. The 54th and 57th harmonics are indicated by other tidal components, but an estimate of noise level may be obtained from the 52nd, 53rd, and 56th, and 58th harmonics. Results of these computations are presented in Figure 2. The phase lag gives the lag of the observed tidal component behind the theoretical equilibrium tide. Because the period of observations is too short to resolve the K_1 and P_1 components, the amplitude cited as (K_1, P_1) applies to a period intermediate to that of K_1 and P_1 , and no phase lag is cited for this component.

The principal differences between Weddell and Ross Sea tides are (1) the Weddell Sea tide may be classed as a mixed tide, having both diurnal and semidiurnal components, whereas the Ross Sea tide is diurnal; (2) the tidal amplitudes are larger in the Weddell Sea than in the Ross Sea, a fact possibly related to the shape of the basins south of the two stations. The Ross Sea tide is unusual in that the diurnal component is predominant.

Tidal Currents. On July 4, 1958, oceanographic studies were made at Kainan Bay near the tip of Antarctica, including measurements with

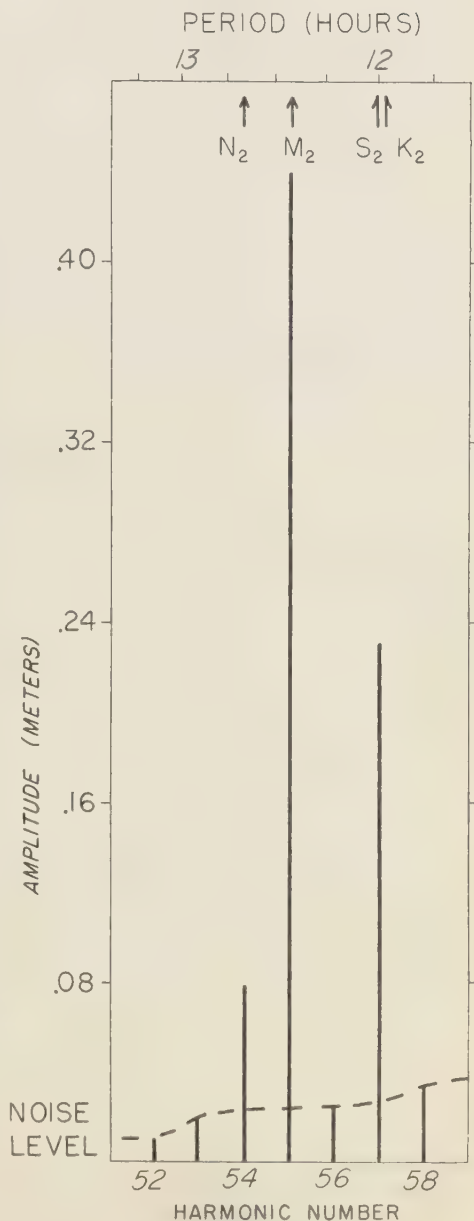


Fig. 7. Computation of amplitude of principal lunar component M_2 and noise level, Weddell Sea tide.

TABLE 2. Amplitude (meters), Noise Level (meters), and Phase Lag of Principal Tidal Components

	M_2	N_2	S_2	O_1	(K_1, P_1)
Weddell Sea tide					
	0.44	0.04	0.23	0.22	0.41
	0.02	0.03	0.04	0.08	0.04
	208.2°	224.5°	239.0°	344.6°	
Ross Sea tide					
	0.03	0.04	0.04	0.21	0.35
	0.02	0.03	0.02	0.06	0.03
	93.0°	21.8°	24.2°	333.8°	

 M_2 = principal lunar component N_2 = larger lunar elliptic component S_2 = principal solar component O_1 = principal lunar diurnal component K_1 = lunisolar diurnal component P_1 = principal solar diurnal component

an Ekman current meter of current during a 6-hour interval at depths of 300 and 400 meters. Simultaneously, gravity measurements were made at Little America V with the hope of obtaining a correlation of tidal currents and vertical shelf movements from which the inward dimension of the shelf might be calculated.

Let

 V_c = northward component of current V_s = rate of change of surface elevation H = thickness of water column between shelf bottom and ocean floor at ice front = 364 meters by seismic measurement L = inward dimension of ice shelf

Then the volume of water flowing out from beneath the shelf per meter of section in a given time ($V_s \times H \times \text{time}$) should equal the product of elevation change at the ice front during this time and the length of floating ice back of

the ice front ($V_c \times L \times \text{time}$). If we plot current velocity against the rate of change of surface elevation at Little America V, we obtain by a least-square fit of the six observations 3.64×10^{-4} m/sec change in surface elevation for 1 m/sec change in current. Thus

$$L = (V_c / V_s) H$$

$$= 364 / (3.64 \times 10^{-4}) = 1,000 \text{ km.}$$

For comparison we note (Fig. 1) that the distance from the ice shelf at Little America V to the southern extremity of the Ross Shelf is 840 km. We may conclude that the gravimetric determination of tidal amplitude of the proper magnitude to account for the observed tidal current.

Acknowledgments. J. McKim Malville, astronomer at Ellsworth, and Hugh F. Bennett, geophysicist at Little America V, participated in the bihourly gravity observations. Lyle D. McGinnis and Stephen den Hartog assisted with the simultaneous gravity and ocean current measurements at Little America V. Patricia L. Griffiths helped with the analysis of the data on the IBM 704 computer. The authors wish to thank the other individuals for their contributions to this study.

REFERENCES

- Baars, B., Gravity effect of earth tides, *Geophysical Prospecting*, 1, 82-110, 1953.
 Crary, A. P., R. D. Cotell, and Jack Oliver, C. Physical studies in the Beaufort Sea, 1955, *Trans. Am. Geophys. Union*, 33, 211-216, 1955.
 Crary, A. P. and Norman Goldstein, Geophysical studies in the Arctic Ocean, *Deep Sea Research*, 4, 185-201, 1957.
 Nettleton, L. L., *Geophysical Prospecting for Oil*, McGraw-Hill, New York, 444 pp., 1940.
 Schureman, P., A manual of the harmonic analysis and prediction of tides, *U. S. Coast and Geodetic Survey Spec. Publ. 98*, 1924.

(Manuscript received October 8, 1959.)

Simplified Method of Determining Refraction Coefficients for Sea Waves

R. DORRESTEIN

*Royal Netherlands Meteorological Institute, Section of Oceanography
De Bilt, The Netherlands*

Abstract. In order to find the effect of wave refraction in one particular locality P only, it is useful to construct wave rays from this point outward to sea and to determine their variation in azimuth from θ at P to θ' in deep water. The change of energy flux density for waves coming from a direction θ' in deep water is, as customary, expressed by the square of a 'refraction coefficient,' K^2 . For waves of a given period whose energy in deep water is uniformly distributed over a finite azimuthal interval from θ_1' to θ_2' , the average square of the refraction coefficient is given (with restrictions) by

$$K^2 = (C'/C) \cdot \{[\theta(\theta_2') - \theta(\theta_1')]/[\theta_2' - \theta_1']\}$$

where C' and C are the phase velocities of the waves in deep water and at P , respectively.

Some practical hints are given. A derivation of the above formula is given in the appendix.

Introduction. The purpose of this note is to draw attention to the possibility of a simplification of the procedure for determining the effect of refraction of waves coming from the sea into shoaling water, especially if we want to know this effect at one particular locality P (for instance, near the end of a proposed breakwater).

In such cases it is convenient to construct wave rays, from the point under consideration, in several directions outward to sea, yielding so-called refraction fan diagram. The usefulness of such fan diagrams was mentioned in a manual of the *Beach Erosion Board* [1954] in which it was deemed necessary (p. 36) to select 'companion orthogonals . . . shoreward on either side of the seaward projected rays in order to determine the refraction coefficient for the various directions of wave approach.' It will be shown that this step can be omitted.

The procedure applied in such cases in this Institute since about 1953, which leads to a simple and straightforward determination of the refraction coefficient K , and which does not seem to have been published, is described below.

Theory. We suppose that we want to know the effect of refraction of the waves coming from the sea to a point P (Fig. 1). As is customary, the laws of 'geometrical optics' are

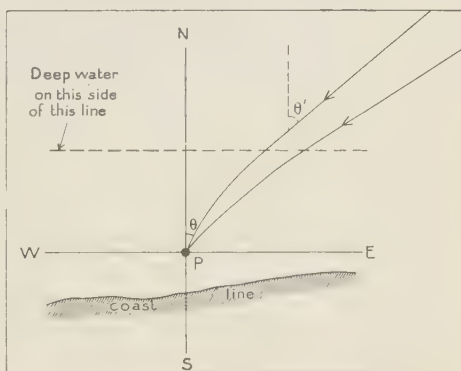


Fig. 1. Two wave rays approaching the point P .

assumed to be applicable. We consider waves of a fixed period T (or a narrow interval of periods). Reflection by the coast or obstacles is neglected.

Notation.

- θ = azimuth of a wave ray at the point P ; may be measured relative to the north or any other convenient direction.
- $\theta'(\theta)$ = azimuth in deep water of the wave ray which has an azimuth θ at the point P .
- C = phase velocity of waves at P .
- C' = phase velocity of waves in deep water.
- V = group velocity of waves at P .

V' = group velocity of waves in deep water.
 g = acceleration of gravity.
 ρ = density of water.

For a field of long-crested waves in deep water arriving from the average direction θ' and passing the point P from the average direction θ :

H = root mean square height at P .
 H' = root mean square height in deep water.
 e = $1/8g\rho H^2 V$ = energy flux density at P
 = amount of wave energy passing, near P , per unit width and per unit time.
 e' = $1/8g\rho H'^2 V'$ = energy flux density in deep water
 = amount of wave energy passing in deep water per unit width and per unit time.
 $K^2(\theta') = e/e' =$ square of refraction coefficient for waves from direction θ' (customary definition).

The formula. In the normal case, a small increment $\delta\theta$ of the angle θ will be accompanied by a proportionally small increment $\delta\theta'$ of θ' . Our procedure, then, is based upon the simple formula

$$K^2(\theta') = (C'/C) |\delta\theta/\delta\theta'| \quad (1)$$

which is derived in the Appendix. The *modulus* of $\delta\theta/\delta\theta'$ is written to cover also the case of a small increase in θ being accompanied by a small decrease of θ' . Cases where a small increment $\delta\theta$ is *not* accompanied by a proportional small increment $\delta\theta'$ will be discussed below.

The procedure. The procedure to be applied for finding the refraction coefficient K for various directions by means of (1) is now obvious. Starting from the point P , wave rays are constructed toward the sea in the directions which seem to be of interest, for instance, with intervals of $\Delta\theta = 10^\circ$. The values found for θ and azimuth in deep water, θ' , are plotted against the θ values, a smooth curve is drawn through the points thus obtained, and from the slope of this curve, the refraction coefficient K is found by means of (1) for any occurring value of θ' .

Singularities. It is only in cases of rather regular bottom topography—for instance, deep contours which are more or less parallel—that θ' can be expected to be a monotonically increasing function of θ . Then (1) can be applied throughout. In cases of somewhat less simple bottom topography, however, other features may be observed.

Figure 2A is a graph of θ' versus θ with $\delta\theta'/\delta\theta = 0$ at points D and F . For a deep wave direction between θ_F' and θ_D' we have more than one corresponding wave direction in P . Such a case occurs, for instance, if rays pass some shoal region which acts as a positive lens. In Figure 2A the points characterizing the rays which are then most influenced by this shoal region lie between D and F . Wave energy approaching in deep water from an interval of directions $\delta\theta_1'$ travels on either side of this shoal and arrives at P from directions in the intervals $\delta\theta_1$ and $\delta\theta_3$, and it also travels

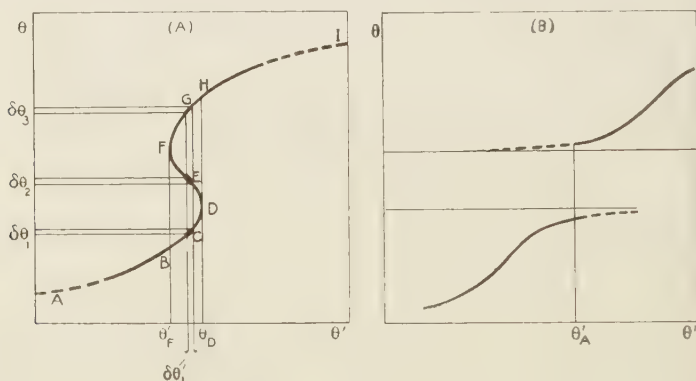


Fig. 2. Two cases where the deep water azimuth θ' is not a monotonic function of the azimuth θ in the point P .

the shoal (no breaking assumed) and then at P from directions in the interval (if it may be assumed that the waves from these directions near P do not disturb each other), we may use formula (1) for the square of the refraction coefficient, K^2 , i.e., the concentration factor for the energy flux, for all the ways of travel (the points C , E , and G in Figure 2A). In the points D and F of Figure 2A, $\partial\theta'/\partial\theta = 0$. This means that P is lying on a caustic line for waves coming from the directions θ_F' and θ_D' in deep water. Applying (1) at these points, K^2 would become infinite. We need to be, however, be alarmed by this because waves at P do not come from one or two directions but from a whole interval. We shall return to this point in the following section.

In general, the slope of the θ' , θ graph, and the value of K^2 , decreases more and more as the direction of the incident waves approaches the coast line or of the depth contours. This is also illustrated near the points A and I in Figure 2A.

Figure 2B a θ' , θ graph is given with two branches separated by an interval of θ values to which no value for θ' corresponds. This is a case that can be due, for instance, to the presence of an island at some distance off the coast which prevents wave rays from arriving at P from a certain range of directions. Figure 2B illustrates the case that waves from certain directions (near θ_A') can reach the point P from both sides of the island, thus leading to a dumbbell wave pattern which is typical for such cases. The application of (1) yields no difficulties here.

Averaging over θ' . The expression (1) for the square of the refraction coefficient K^2 lends itself especially well to integration or averaging over the deep water azimuth θ' . The average value $\overline{K^2}$ for waves of one period (or a narrow interval of periods) whose energy in deep water is uniformly distributed over a finite interval θ_1' to θ_2' is apparently

$$\overline{K^2} = \frac{C'}{C} \left| \frac{\theta(\theta_2') - \theta(\theta_1')}{\theta_2' - \theta_1'} \right|. \quad (2)$$

It can be seen that this expression is valid not only if $\theta'(\theta)$ and $\theta(\theta')$ are monotonic functions over the intervals considered, but also in a case as

illustrated in Figure 2A, provided the interval from θ_1' to θ_2' includes the interval from θ_F' to θ_D' . In other cases the integration is also easy but some care should be taken to avoid errors.

In principle, of course, for real sea waves with a given main direction in deep water but characterized by a two-dimensional energy spectrum (distribution of the energy over the directions as well as over the periods or wave numbers), the wave energy at a point P should be found by integrating the energy contributions at P due to all directions and wave numbers involved. If the energy spectrum in deep water is known, this can be done, but it will be very laborious. The averaging of K^2 over a finite θ' interval can be considered as a first step in this direction, and since this step is so simple it is desirable to carry it out throughout and thus to use (2) rather than (1) in all practical cases.

The question then arises as to the magnitude of the θ' interval to be chosen. In practical applications, we took 40° for this interval; for typical swell waves a smaller value is better. Since in cases as illustrated by Figure 2A, the interval $\theta_D' - \theta_F'$ rarely exceeds 10 to 20° , the complications discussed earlier are greatly reduced with this method of averaging.

Practical execution of ray construction. The procedure to be applied for finding the refraction coefficient K was indicated above. Since it is especially the variations of azimuth along the rays starting from the selected point P which

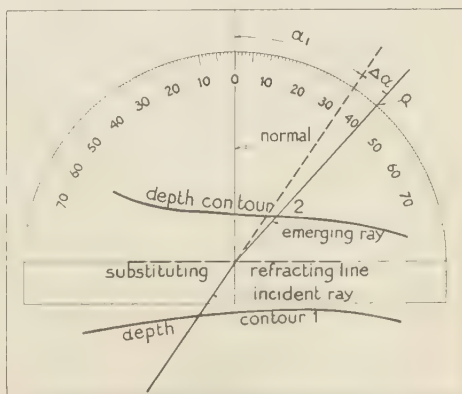


Fig. 3. Illustration of the use of semicircular transparent scale.

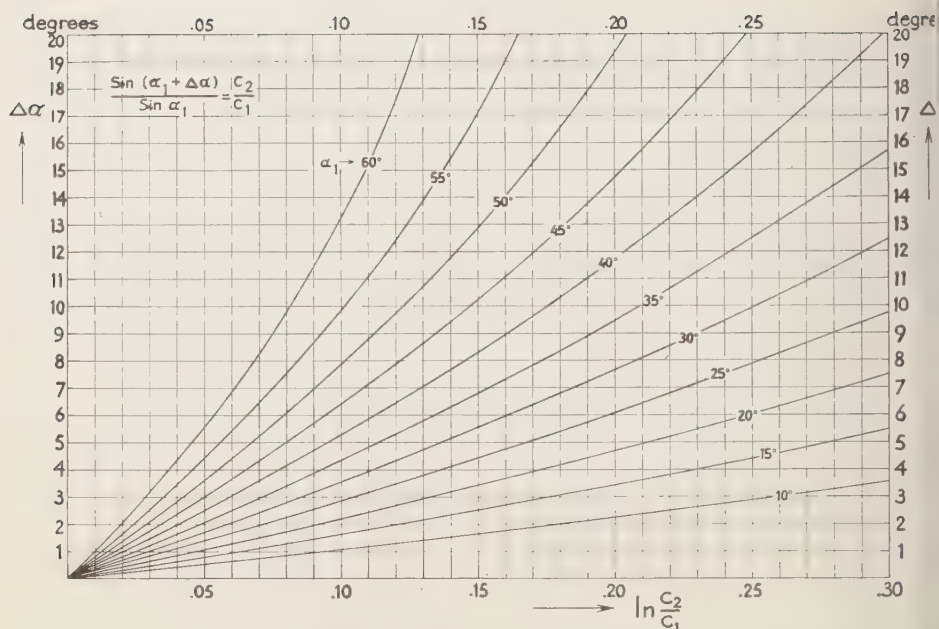


Fig. 4. Auxiliary diagram for reading the variation of azimuth, $\Delta\alpha = \Delta\theta$, by one refraction-step.

are of interest, we found the following method for determining the function $\theta'(\theta)$ most useful.

As usual, the refraction is approximated by a number of finite steps. The continuous refraction in the region between two depth contours, 1 and 2, with wave velocities C_1 and C_2 (in general $C_2 > C_1$), is approximated by a discontinuous refraction at an estimated intermediate contour. To determine the effect of refraction between two depth contours: (1) a semicircular transparent protractor is laid with its axis along the intermediate contour substituting the region between depth contours 1 and 2 (see Fig. 3: 'substituting refracting line'); (2) the angle α_1 between the incident ray and the normal to this mean contour is read in whole degrees and noted in a small auxiliary table; (3) the variation of azimuth, $\Delta\alpha = \Delta\theta$, is read in tenths of degrees on an auxiliary diagram (Fig. 4) and noted; (4) $\Delta\alpha$ is added (mentally) to α_1 to give the new angle α_2 , and a point is set (Q in Fig. 3) through which the continuation of the ray must pass; (5) the transparent scale is removed and the continuation of the ray is drawn by pencil.

After repeated application until the ray in deep water, the successively noted values $\Delta\alpha$ are added to θ to yield θ' . If the angle exceeds 60 to 70° other methods must be used as described in the literature [for instance Arthur, Munk, and Isaacs, 1952].

In Figure 4, the choice $f(C_2/C_1) \equiv \ln(C_2/C_1)$ for the independent variable, or for the 'refracting power' of the region between depth contours 1 and 2, has been made because of additive property: for three depth contours 2, and 3 we have

$$f(C_3/C_1) = f(C_2/C_1) + f(C_3/C_2)$$

so that successive steps may be easily combined if the depth contours are nearly parallel.

APPENDIX

The relation (1) can be considered to be nothing but a direct consequence of an general geometrical-optical law which was first published by Straubel [1903] but which essentially seems to be contained in the work of R. Clausen [see Tiddens, 1904; Boeghold, 1924; Maréchal, 1924].

$$K^2 = (C'/C)(\delta\theta/\delta\theta''). \quad (A3)$$

This is almost identical with the relation to be proved. Remembering that all variables with δ are meant to represent infinitesimally small quantities, we have still to prove that $\delta\theta''/\delta\theta'$ tends to unity as $\delta\theta'$ is made to approach zero, with the point P and the ray (1) being fixed, and the ray (2) remaining parallel with ray (1). This is most easily demonstrated by first keeping the point Q fixed with respect to P but letting the points P' and Q' shift to infinity. If we think of rays (1) and (2) as being fixed and denote the points where ray (3) through P and ray (4) through Q cross a line AB (Fig. 5) by A^* and B^* , respectively, we can conclude from the continuity of the refraction that AA^* and $\delta\theta'$ are continuous functions of $\delta\theta$ and that BB^* and $\delta\theta''$ are continuous functions of $\delta\theta'$. This means that AA^* and $\delta\theta'$ tend to zero as $\delta\theta$ approaches zero, and that BB^* and $\delta\theta''$ tend to zero as $\delta\theta'$ approaches zero. We denote, as before, the intersection point of rays (2) and (3) at Q' and the intersection point of rays (1) and (4) at P' and allow both $\delta\theta$ and $\delta\theta'$ to approach zero in such a way that $P'Q'$ remains perpendicular to rays (1) and (2) so that the relations (A1) and (A3) remain valid. Denoting the distances $AP' = BQ'$ by F , we then see that both $F/\delta\theta' = B/1^*$ and $F/\delta\theta'' = AB^*$ tend to the same value AB ; thus $\delta\theta''/\delta\theta'$ tends to unity and (A3) can be replaced by

$$K^2 = (C'/C) |\delta\theta/\delta\theta'| \quad (1)$$

This is the relation to be proved.

For the sake of completeness we may add that the relation (1) can also be derived in a rather indirect, but instructive, way from Liouville's theorem (which is a fundamental law in classical statistical mechanics), by using the generally valid particle-ray analogy [see,

e.g., Eckart, 1948, 1950]. Liouville's theorem states that the density of a continuous cloud of points in the phase space remains constant as we follow any point. The ray-analogy states that the spectral density of waves in the number space remains constant as we follow any wave group. The relation recently derived from geometrical-optical considerations by Longuet-Higgins [1957] is no more than this ray-analogue of Liouville's theorem. Our relation can be deduced from his equations (8) and

Acknowledgment. The Director-in-Chief of the Royal Netherlands Meteorological Institute granted me his permission to publish this paper.

REFERENCES

- Arthur, R. S., W. H. Munk, and J. D. Isaacs, The direct construction of wave rays, *Trans. Geophys. Union*, 33, 855-865, 1952.
- Beach Erosion Board, Shore protection planning and design, *Tech. Rept. 4*, Corps of Engineers, Washington, D. C., 1954.
- Boegehold, H., Die allgemeinen Gesetze über Lichtstrahlenbündel usw., in *Grundzüge der Theorie der optischen Instrumente*, edited by Czapski-Eppenstein, 3. Auflage, Barth, Leipzig, pp. 213-233, 1924.
- Eckart, C., The approximate solution of three-dimensional wave equations, *Revs. Mod. Phys.*, 20, 399-417, 1948.
- Eckart, C., The ray-particle analogy, *J. Math. Research, Sears Foundation*, 9, 139-144, 1955.
- Longuet-Higgins, M. S., On the transformation of a continuous spectrum by refraction, *Proc. Cambridge Phil. Soc.*, 53, 226-229, 1957.
- Maréchal, A., Optique géométrique générale, *Encyclopedia of Physics*, edited by S. Flügge, Berlin, Göttingen, Heidelberg, p. 56, 1956.
- Straubel, R., Über einen allgemeinen Satz der geometrischen Optik und einige Anwendungen, *Phys. Z.*, 4, 114-117, 1903.
- Tiddens, P. G., Beschouwingen over den loop der lichtstralen en de beeldvorming in optische stelsels, Thesis, Univ. of Leiden, 97 pp., 1959.

(Manuscript received August 31, 1959.)

A New Approach to Peak Flow Estimation

R. RANGARAJAN

Central Water and Power Commission, New Delhi, India

Abstract. The usual flow prediction methods so far advanced proceed by considering each year's data as constituting a sample. They do not take due account of the nonrandom nature of the daily flows in a year. This paper outlines a different method in which less restrictive assumptions are made and proper account of the seasonal and intercorrelation effects is taken. Actual application of the method is illustrated by an example.

Introduction. Accurate assessment of the maximum daily flow in a river with a sufficiently long period of recurrence has been a fundamental problem for irrigation engineers engaged in planning storage reservoirs and other water-utilization measures. This being basically a prediction problem, the experience of the past should be sought to bear upon the methods employed. Unfortunately, the data are scanty, and the available information about the past is not known. However, under certain plausible assumptions about the nature of the data, a number of flood-prediction methods based upon probability theory have been advanced at various times during the course of the past two decades. In these methods each year's data are usually considered as constituting a sample, and from a number of such sample years' data appropriate forecasting formulas are evolved. But the methods do not take due account of the nonrandom nature of the daily flows in a year and the tendencies of these high flows to cluster particularly in tropical rivers, as in India, where flows are conditioned by monsoons. Even the popular method of *Gumbel* [1954], who utilizes the distribution theory of extreme values, suffers from a similar limitation. Accordingly, a new approach, which has less restrictive assumptions and which takes proper account of the seasonal and inter-correlation effects, is explored below.

The method. For this purpose, the problem is to be reoriented and viewed from a different angle. Instead of considering the data as yearly samples, as is implied in most of the frequency methods sponsored hitherto, it will be advantageous to visualize the discharge values observed over a number of years on any specified

date, like June 1, June 2, as giving rise to a frequency distribution. Thus one can obtain a number of such frequency distributions, each corresponding to a particular date of the year. But these frequency distributions on the different dates are not quite independent of each other, for the discharge values on any date are influenced by preceding dates' records. It is our object to determine the probability that any assigned high flood will occur at least once among the different dates of the year. This has to be obtained from the frequency distribution of the individual dates and their intercorrelations. Let us now develop the necessary theory for this purpose.

Let $x_1, x_2, \dots, x_1, \dots, x_n$ be n variates ordered according to time, such as the discharges on n consecutive dates of the season. Our problem is to find the probability that any measure X will be exceeded by at least one of the variates x_1, x_2, \dots, x_n ; and this is just the complement of the probability that none of the values on different dates exceeds the assigned X . The latter probability may be written as

$$Pr(x_1 \leq X, \quad x_2 \leq X, \dots, x_n \leq X)$$

for the evaluation of which we stipulate certain conditions.

The x_i 's have the same distribution function, which is known; each x_i is correlated only with x_{i-1} ; and the joint distribution of any two adjacent pairs of x 's is the same, irrespective of the dates. These can be written in algebraic notation as follows:

1. $F_i(X) = Pr(x_i \leq X)$ is defined and is the same for all i 's = $F(X)$.

2. $F_{i,i-1}(XY) = Pr(x_i \leq X, x_{i-1} \leq Y)$ is known and is the same for all i 's = $F(XY)$.
3. $Pr(x_i \leq X/x_{i-1}, x_{i-2} \dots) = Pr(x_i \leq X/x_{i-1})$ for all i 's.

Using condition 3 first, we derive

$$\begin{aligned}
 & Pr(x_1 \leq X, x_2 \leq X, \dots, x_n \leq X) \\
 &= Pr(x_1 \leq X, \dots, x_{n-1} \leq X) \\
 &\quad \cdot Pr(x_n \leq X/x_1 \leq X, x_2 \\
 &\quad \leq X, \dots, x_{n-1} \leq X) \\
 &= Pr(x_1 \leq X, x_2 \leq X, \dots, x_{n-1} \leq X) \\
 &\quad \cdot Pr(x_n \leq X/x_{n-1} \leq X) \\
 &= Pr(x_1 \leq X, x_2 \leq X, \dots, x_{n-1} \leq X) \\
 &\quad \cdot \frac{Pr(x_{n-1} \leq X, x_n \leq X)}{Pr(x_{n-1} \leq X)} \\
 &= Pr(x_1 \leq X, x_2 \leq X, \dots, x_{n-1} \leq X) \\
 &\quad \cdot F_{n-1,n}(X, X)/F_{n-1}(X)
 \end{aligned}$$

that is, $Pr(x_1 \leq X, \dots, x_n \leq X)$

$$= \frac{F_{12}(XX)F_{23}(XX) \dots F_{n-1,n}(XX)}{F_2(X)F_3(X) \dots F_{n-1}(X)} \quad (1)$$

More generally,

$$\begin{aligned}
 & Pr(x_1 \leq X_1, x_2 \leq X_2, \dots, x_n \leq X_n) \\
 &= \frac{F_{12}(X_1X_2)F_{23}(X_2X_3) \dots F_{n-1,n}(X_{n-1}X_n)}{F_2(X_2)F_3(X_3) \dots F_{n-1}(X_{n-1})} \quad (2)
 \end{aligned}$$

Now if we apply the first two conditions, the two expressions can be considerably simplified into

$$Pr(x_1 \leq X \dots x_n \leq X) = \frac{[F(XX)]^{n-1}}{[F(X)]^{n-2}} \quad (3)$$

and

$$\begin{aligned}
 & Pr(x_1 \leq X_1 \dots x_n \leq X_n) \\
 &= \frac{F(X_1X_2)F(X_2X_3) \dots F(X_{n-1}X_n)}{F(X_2)F(X_3)F(X_4) \dots F(X_{n-1})} \quad (4)
 \end{aligned}$$

It is seen that though very restrictive conditions have been imposed in order to simplify

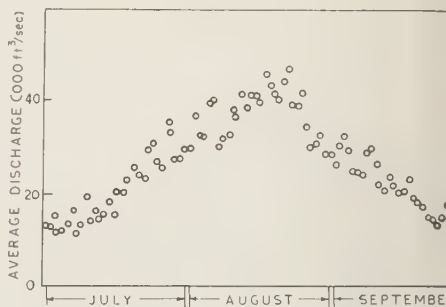


Fig. 1. Yamuna River at Tajewala. Trend the average discharges on different dates, averaged over 29 years (1927-55).

the final expression, conditions 1 and 2 can be relaxed sufficiently to permit us to apply the method to many general practical cases, as the example given below discloses. The chief merits of the method lie in the fact that it makes use of the basic data at our disposal and also takes proper account of the intercorrelations between successive days' records.

Application of the method. As an illustration of its scope, the method was applied to 29 years' daily flow records available for the Yamuna River at Tajewala, about 100 miles from Delhi, and the results are discussed. The peak floods in this river normally occur in the months of July, August, and September. Hence it was considered sufficient to restrict the analysis to only these three months' daily flow data. At the outset, it was necessary to examine how far the basic requirements or conditions were satisfied by the data. In the first place, as a result of the seasonal effects, the average runoff over 29 years on various dates plotted in Figure 1 reveals an upward trend from July to August 20 followed later on by a more or less steady recession. This meant that the frequency distributions of the flows on the various dates did not have the same mean position. This changing position of the mean flow on these dates only pointed to what was known in statistical parlance as the non-stationarity of the different dates' flow patterns. This feature hence had to be taken into consideration in the application of the method. This was found possible, though it entailed an involved procedure.

For determining the form of the distribution

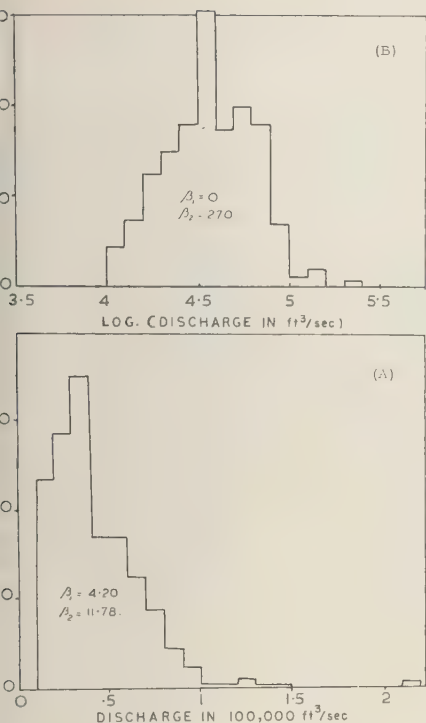


Fig. 2. (a) Distribution pattern of daily Yamuna discharges, Aug. 14-24; (b) Distribution pattern of logarithms of daily Yamuna discharges, Aug. 14-24.

For each date's flow records follow, the data for the mid-runoff period from August 14 to 24, during which the means were found in Figure 1 to remain fairly steady, were examined by plotting their frequency histograms both for linear intervals and logarithmic intervals in Figures 2(a) and 2(b). It is clear that the latter were more symmetrical and nearer to gaussian distribution than the former. For the discharge frequency distribution, the skewness (β_1 , indicating the degree of asymmetry) and the kurtosis (β_2 , measuring the flatness of peak) were found to be 4.2 and 11.8, whereas these parameters for the frequency distribution of logarithms for the same discharge values were 0 and 2.7, respectively. These indicated that the individual date's distribution could be taken as following log-normal pattern.

Hence, taking the logarithms of discharges

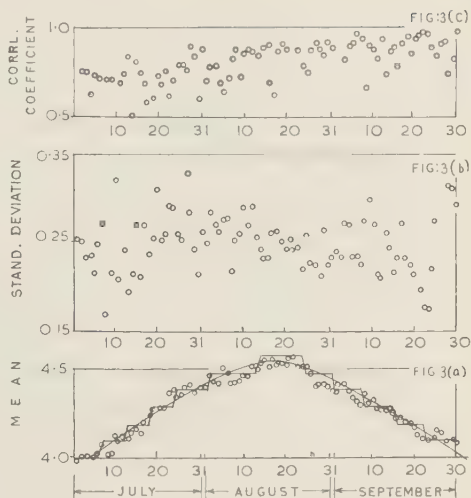


Fig. 3. Yamuna at Tajewala (1927-55). Means, standard deviations and correlation coefficients of logarithms of discharges on different dates.

as the variates, their means and standard deviations for the 92 dates have been plotted in Figures 3(a) and 3(b). Figure 3(c) shows the correlation coefficients of each date's log discharge values with the preceding date's ones. The correlation coefficients varied from 0.5 to 0.997. A small rising trend observed during the year was, of course, not significant compared with the scatter of the points. Hence it would be quite safe to assume an average correlation of 0.85. It may be stated in this connection that the partial correlation between each date's discharge with that observed two days earlier was found to be insignificant, and hence condition 3 in our method was satisfied by the data. The standard deviations in Figure 3(b) were practically random and did not reveal any kind of structure. Hence it was possible to assume an average standard deviation of 0.26 for the distribution on each of the dates. To take due cognizance of the trend in the means, as disclosed in Figure 3(a), a smooth curve was drawn, from which the mean values rounded off to the first decimals were read.

Thus sets of data corresponding to different dates were obtained; each set had the same standard deviation but the means differed from set to set. Also data for any two consecutive

TABLE 1. Standardized Values of Mean Log (discharge)

Date	July			August			September		
	Mean Discharge, ft ³ /sec $\times 10^3$	Mean Log (discharge)	Standardized Value of Mean Log (discharge)	Mean Discharge, ft ³ /sec $\times 10^3$	Mean Log (discharge)	Standardized Value of Mean Log (discharge)	Mean Discharge, ft ³ /sec $\times 10^3$	Mean Log (discharge)	Standardized Value of Mean Log (discharge)
1	12.35	4.0	5.1	30.06	4.4	3.5	26.57	4.4	3.5
2	12.26	4.0	5.1	37.34	4.5	3.1	30.60	4.4	3.5
3	15.08	4.0	5.1	32.87	4.5	3.1	33.11	4.4	3.5
4	11.64	4.0	5.1	32.66	4.5	3.1	29.94	4.4	3.5
5	11.68	4.0	5.1	39.66	4.5	3.1	25.63	4.4	3.5
6	13.24	4.0	5.1	40.75	4.5	3.1	24.99	4.4	3.5
7	16.12	4.1	4.7	30.25	4.5	3.1	24.54	4.4	3.5
8	11.52	4.1	4.7	32.27	4.5	3.1	29.50	4.3	3.9
9	12.87	4.1	4.7	33.23	4.5	3.1	30.08	4.3	3.9
10	19.64	4.1	4.7	38.67	4.5	3.1	26.74	4.3	3.9
11	13.81	4.1	4.7	36.83	4.5	3.1	22.44	4.3	3.9
12	16.14	4.1	4.7	41.98	4.5	3.1	20.87	4.3	3.9
13	14.33	4.2	4.3	38.89	4.5	3.1	24.21	4.3	3.9
14	15.41	4.2	4.3	41.76	4.6	2.7	22.17	4.3	3.9
15	18.17	4.2	4.3	41.88	4.6	2.7	20.84	4.3	3.9
16	15.55	4.2	4.3	40.01	4.6	2.7	20.84	4.2	4.3
17	20.45	4.2	4.3	46.57	4.6	2.7	23.72	4.2	4.3
18	20.43	4.2	4.3	43.79	4.6	2.7	19.52	4.2	4.3
19	23.47	4.3	3.9	42.15	4.6	2.7	18.57	4.2	4.3
20	25.75	4.3	3.9	41.39	4.6	2.7	17.36	4.2	4.3
21	24.14	4.3	3.9	45.20	4.6	2.7	15.59	4.2	4.3
22	23.90	4.3	3.9	47.79	4.6	2.7	14.34	4.2	4.3
23	29.82	4.3	3.9	39.94	4.6	2.7	13.69	4.1	4.7
24	30.97	4.3	3.9	39.91	4.6	2.7	15.13	4.1	4.7
25	27.10	4.4	3.5	42.29	4.5	3.1	18.10	4.1	4.7
26	26.82	4.4	3.5	34.57	4.5	3.1	32.18	4.1	4.7
27	35.64	4.4	3.5	30.33	4.5	3.1	29.80	4.1	4.7
28	33.44	4.4	3.5	30.85	4.5	3.1	19.84	4.1	4.7
29	27.78	4.4	3.5	32.94	4.5	3.1	18.54	4.1	4.7
30	28.19	4.4	3.5	29.16	4.5	3.1	16.75	4.1	4.7
31	29.78	4.4	3.5	28.97	4.5	3.1			

dates had the same correlation coefficient. The consecutive dates having the same mean on account of rounding off to the first decimal are grouped by bars in Figure 3(a). Any two consecutive dates' discharges are thus generally distributed as bivariate log-normal distribution with different or sometimes the same means but always with the same standard deviation, 0.26, and correlation coefficient, 0.85. Because of the variation of the means, the distribution functions though having the same form differed in their locations. Hence the more general formula given by (2) had to be employed.

For example, to calculate the probability that

all 92 values of the season would be ≤ 200 , ft³/sec, its logarithm ($= 5.3010$) should first be expressed in standardized form for each date, using that date's mean and standard deviation. For July 1, the mean of the discharge was 4.0 (rounded to first decimal) and its standardized value was equal to $(5.3010 - 4.0)/0.2557$ or 5.1. Similarly, the standardized values for each of the dates have been given in Table 1.

It was then necessary to calculate the functions $F_i(X_i)$ and $F_{i-1,i}(X_{i-1}, X_i)$. The first of functions were quite easily obtained from the usual normal probability tables. On

TABLE 2. Extension of Bivariate Normal Table for $x > 3.1$, $y > 3.1$, and $\rho = 0.85$

y	$u = \frac{3.1 - \rho y}{\sqrt{1 - \rho^2}}$	Ordinate $= \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{1}{2}y^2)$	Area $= 1 - \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} \cdot \exp(-\frac{1}{2}u^2) du$
3.1	0.8826	0 ² 32668	0.1887286
3.2	0.7213	0 ² 3841	0.2353637
3.3	0.5599	0 ² 17226	0.2877739
3.4	0.3986	0 ² 12322	0.3450949
3.5	0.2372	0 ² 87270	0.4062517
3.6	0.0758	0 ² 61190	0.4697894
3.7	-0.0855	0 ² 42480	0.5340677
3.8	-0.2469	0 ² 29190	0.5975062
3.9	-0.4082	0 ² 19870	0.6584354
4.0	-0.5696	0 ² 13380	0.7155252
4.1	-0.7310	0 ² 89300	0.7676094
4.2	-0.8923	0 ² 58900	0.8138818
4.3	-1.0537	0 ² 38500	0.8539870
4.4	-1.2150	0 ² 24900	0.8878141
4.5	-1.3764	0 ² 16000	0.9156486
4.6	-1.5378	0 ² 10141	0.9379496
4.7	-1.6991	0 ² 63698	0.9553491
4.8	-1.8605	0 ² 39613	0.9685922
4.9	-2.0218	0 ² 24390	0.9784007
5.0	-2.1832	0 ² 14867	0.9854886
5.1	-2.3446	0 ² 8972	0.9904755
5.2	-2.5059	0 ² 5361	0.9938924
5.3	-2.6673	0 ² 3171	0.9961766
5.4	-2.8286	0 ² 1857	0.9976623
5.5	-2.9900	0 ² 1077	0.9986051
5.6	-3.1514	0 ² 618	0.9991875
5.7	-3.3127	0 ² 351	0.9995379
5.8	-3.4741	0 ² 198	0.9997437
5.9	-3.6354	0 ² 110	0.9998612

$$\begin{aligned}
 \text{Volume} &= \frac{1}{\sqrt{2\pi}} \int_{3.1}^{\infty} \exp(-\tfrac{1}{2}y^2) dy \int_u^{\infty} \exp(-\tfrac{1}{2}u^2)/\sqrt{2\pi} du \\
 &= \int_{3.1}^{\infty} dy \left\{ [\exp(-\tfrac{1}{2}y^2)/\sqrt{2\pi}] \left(\int_u^{\infty} \exp(-\tfrac{1}{2}u^2)/\sqrt{2\pi} du \right) \right\} \\
 &= 0.00032676 \quad \text{By Simpson's rule.}
 \end{aligned}$$

her hand, the second set of functions were t all obtainable from Pearson's bivariate rmal tables, as the tables extended only up $F(2.6, 2.6)$. Hence they were calculated, ing the univariate normal values, by finite egration. Table 2 illustrates the calculations ade in this regard for the volume of bivariate rmal distribution for $x > 3.1$ and $y > 3.1$ hen the correlation between the variates is 35. This volume is

$$\frac{1}{\sqrt{2\pi}} \int_{3.1}^{\infty} \exp(-\tfrac{1}{2}y^2) dy$$

$$\begin{aligned}
 &\cdot \int_u^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}u^2) du, \quad \text{where } u \\
 &= \frac{x - \rho y}{\sqrt{1 - \rho^2}}, \quad \rho = 0.85 \\
 &= \int_{3.1}^{\infty} \left\{ \left(\frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}y^2) \right) \right. \\
 &\quad \cdot \left. \int_u^{\infty} \left(\frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}u^2) du \right) \right\} dy
 \end{aligned}$$

TABLE 3. Estimation of Return Period for Peak Flood Discharge of 200,000 ft³/sec in Yamuna River [log (200,000) = 5.3010, σ = 0.2557]

Dates	Means of Log Discharge for Consecutive Pairs of Days	No. of Pairs	Standardized Values of Log (200,000) Corresponding to Means of Pairs in Col. 2, $x_{i-1} - x_i$	$F'_{i-1,t} \cdot (X_{i-1}, X_i)$	$F_{i-1} \cdot (X_{i-1})$	$F_i(X_i)$	$F_{i-1,t} \cdot (X_{i-1}, X_i)$	Log (Col. 6)	Log (Col. 7)
1	2	3	4	5	6	7	8	9	10
				.00	.99	.99	.99	-.00	-.00
July 1-6	4.0-4.0	5	5.1-5.1	000002	999983	999983	999968	000006	000001
6-7	4.0-4.1	1	5.1-4.7	000006	999983	999870	999859	000006	000006
7-12	4.1-4.1	5	4.7-4.7	000021	999870	999870	999761	000060	000100
12-13	4.1-4.2	1	4.7-4.3	000052	999870	999150	999072	000060	000399
13-18	4.2-4.2	5	4.3-4.3	000170	999150	999150	998470	000368	000677
18-19	4.2-4.3	1	4.3-3.9	000384	999150	995190	994724	000368	002229
19-24	4.3-4.3	5	3.9-3.9	001156	995190	995190	991536	002091	003677
24-25	4.3-4.4	1	3.9-3.5	002422	995190	976740	974352	002091	011133
July 25-Aug. 1	4.4-4.4	7	3.5-3.5	006665	976740	976740	960145	010103	017322
Aug. 1-2	4.4-4.5	1	3.5-3.1	012983	976740	903240	892963	010103	046511
2-13	4.5-4.5	11	3.1-3.1	032676	903240	903240	839156	042043	069911
13-14	4.5-4.6	1	3.1-2.7	059269	903240	853300	615809	042043	167188
14-24	4.6-4.6	10	2.7-2.7	136472	653300	653300	443072	150830	242544
24-25	4.6-4.5	1	2.7-3.1	059269	653300	903240	615809	150830	167188
25-31	4.5-4.5	6	3.1-3.1	032676	903240	903240	839156	042043	069911
Aug. 31-Sept. 1	4.5-4.4	1	3.1-3.5	012983	903240	976740	892963	042043	046511
Sept. 1-7	4.4-4.4	6	3.5-3.5	006665	976740	976740	960145	010103	017322
7-8	4.4-4.3	1	3.5-3.9	002422	976740	995190	974352	010103	011133
8-15	4.3-4.3	7	3.9-3.9	001156	995190	995190	991536	002091	003677
15-16	4.3-4.2	1	3.9-4.3	000384	995190	999150	994724	002091	002229
16-22	4.2-4.2	6	4.3-4.3	000170	999150	999150	998470	000368	000677
22-23	4.2-4.1	1	4.3-4.7	000052	999150	999870	999072	000368	000399
23-30	4.1-4.1	7	4.7-4.7	000021	999870	999870	999751	000060	000100

$$Pr(x_1 \leq X_1, x_2 \leq X_2, \dots, x_n \leq X_n) = \frac{F_{12}(X_1 X_2) F_{23}(X_2 X_3) \dots F_{n-1,n}(X_{n-1} X_n)}{F_2(X_2) F_3(X_3) \dots F_{n-1}(X_{n-1})}; \text{ In the present case, log (Numerator)} \\ \bar{1} \cdot 95652897; \text{ log (Denominator)} = \bar{1} \cdot 97355640; \text{ log } (Pr) = \bar{1} \cdot 98297257; Pr(x_1 \leq X_1, x_2 \leq X_2, \dots, x_n \leq X_n) = 0 \cdot 961551 \\ \therefore \text{ Return Period} = \frac{1}{1 - Pr} = 26 \text{ years.}$$

Column 3 of Table 2 gives the values of exp $(-\frac{1}{2}y^2)/\sqrt{2\pi}$ for values of $y \geq 3.1$ in steps of 0.1. Corresponding to each value of y , the area over $x \geq 3.1$ or its corresponding standardized value of $u = (3.1 - \rho y)/\sqrt{1 - \rho^2}$, $\rho = 0.85$ has been obtained from the univariate normal tables and tabulated in column 4. By applying Simpson's rule to the products of the series of values in columns 3 and 4, the required volume $x \geq 3.1$ was obtained as 0.00032676. The volume so obtained has been denoted by $F_{12}^1(3.1, 3.1)$, from which the required function $F_{12}(3.1, 3.1)$ could be derived by the relation

$$F_{12}(3.1, 3.1) = F_{12}^1(3.1, 3.1) + F_1(3.1) + F_2(3.1) - 1 \quad (5)$$

where

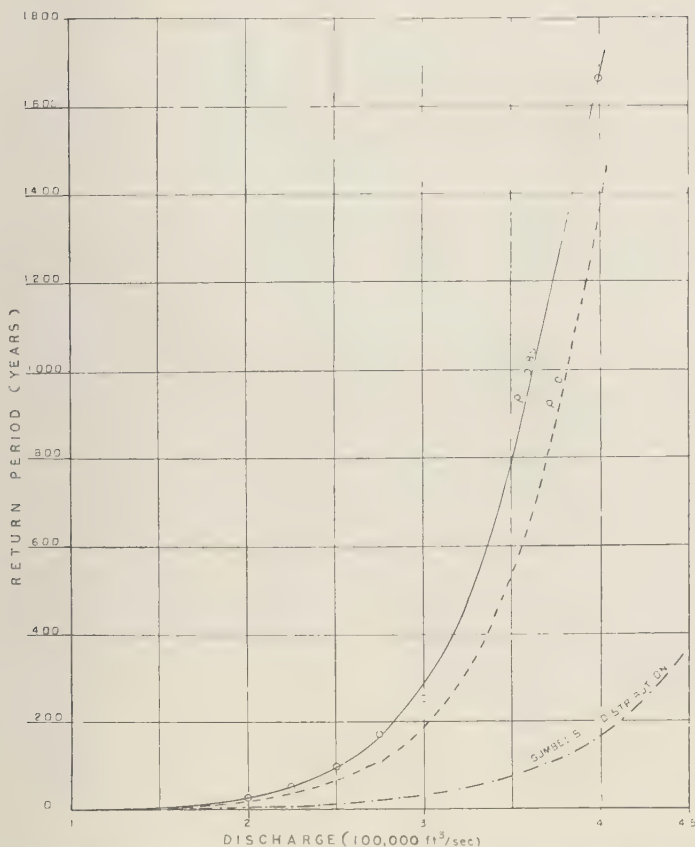
$$F_{12}(3.1, 3.1) = \text{Prob. } (x_1 \leq 3.1, x_2 \leq 3.1)$$

$$F_{12}^1(3.1, 3.1) = \text{Prob. } (x_1 > 3.1, x_2 > 3.1)$$

$$F_1(3.1) = \text{Prob. } (x_1 \leq 3.1)$$

$$F_2(3.1) = \text{Prob. } (x_2 \leq 3.1)$$

The values of similar functions, as required by formula (2) for various dates, were similarly worked out and are compactly given in Table 3 from which the return period for a single-day flood discharge of 200,000 ft³/sec to occur at least once in the season has been estimated as 26 years. On the same lines, the return period for flood discharges 100,000, 200,000, 225,000



4. Yamuna River at Tajewala. Comparison of return period curves for different magnitudes of daily flow by the new approach with $\rho = 0.85$ and $\rho = 0$ and by the method of Gumbel.

000, 275,000, 300,000, and 400,000 ft³/sec were also been evaluated and plotted in Figure 4 a smooth curve drawn through them. For purpose of comparison, similar curves obtained when the correlation between the adjacent values was assumed to be zero, and also by Gumbel's method, were also drawn in the same figure. It is found that the new method yields a steeper curve than the Gumbel procedure. Table 4 gives the daily flow estimates corresponding to typical return periods 20, 50, and 200 years by the two methods, as read from the curves in Figure 4.

Discussion of the results. It is found from Table 4 that when proper account of the correlation effects of successive pairs of daily

flow values is taken, the new flood estimates are about 30 per cent less than the estimates obtained by Gumbel's method. Even when the correlation between successive pairs is assumed to be zero, the new method gives estimates that are 18 per cent less than Gumbel's. In fact, formula (4), which was employed in the present analysis, reduces to

$$\begin{aligned} Pr(x_1 \leq X_1, x_2 \leq X_2, \dots, x_n \leq X_n) \\ = F(X_1)F(X_2) \dots F(X_n) \end{aligned}$$

when the correlation between pairs is assumed to be zero. This is the expression for the exact distribution of extreme values, and its estimates are hence comparable with those obtained by

TABLE 4. Estimates of Daily Flow

Return period, years	Estimated Peak Flood Discharges in 100,000 ft ³ /sec			
	Gumbel's distribution using			
	$\rho = 0.85$	$\rho = 0$	Annual Max Mean = 1.33 Std deviation = 0.83	Annual Max Mode = 0.92 Mean deviation = 0.59
20	1.85	2.05	2.88	2.72
50	2.25	2.40	3.49	3.28
100	2.55	2.70	3.94	3.70
200	2.85	3.05	4.39	4.12

Gumbel's method. The discrepancy appears to be due to the ignoring of the non-stationarity (defined before) of different dates' flow distributions caused by the seasonal variations. In other words, the Gumbel method and other frequency procedures imply that the chance of occurrence of any specific flood magnitude is the same on any day of the year. The present procedure, however, takes due cognizance of these aspects and hence is expected to yield more precise estimates.

Acknowledgments. The author wishes to thank Shri R. Narayanan of the Central Water and

Power Research Station for assisting in the laborious computations, such as the extension of the bivariate normal tables.

REFERENCES

- Feller, W., *An Introduction to Probability Theory and Its Application*, John Wiley & Sons, New York, 1957.
 Gumbel, E. J., Statistical theory of extreme values and some practical applications, *Natl. B. Standards Appl. Math. Ser.*, no. 33, 1954.

(Manuscript received March 20, 1959;
 revised September 18, 1959.)

Runoff as a Function of Moisture-Storage Capacity

J. L. THAMES AND S. J. URSIC

*Southern Forest Experiment Station
Forest Service, U. S. Department of Agriculture
Oxford, Mississippi*

Abstract. Records from small watersheds in northern Mississippi indicate that surface runoff is strongly correlated with storage opportunity in the upper 6 inches of soil. A logarithmic expression was developed to describe these relationships. A procedure for calculating antecedent soil-moisture storage over a watershed is presented.

Introduction. *Whelan and others* [1952] made successful estimates of runoff by using an accounting procedure to route storm water through the profile. The calculations were based largely on pore volume and percolation rates determined from soil samples. By applying a retention index derived from estimates of antecedent storage space in the upper 3 feet of the profile, *Hartman and others* [1958] were able to estimate runoff for a 3-acre watershed in the blacklands of Texas. Examination of soil-moisture and runoff data from several watersheds in northern Mississippi indicated that the amount of water held in the 0- to 6-inch layer at the time of rain affected infiltration and also provided an index to the storage capacity of the soil profile. Furthermore, there appeared to be good correlation between the amount of runoff and the antecedent storage in this layer. A study was therefore undertaken to learn whether these apparent relations of moisture-storage and rainfall-runoff could be used to estimate storm runoff.

Study areas. Watersheds II and III, from a series of three abandoned-field watersheds described by *Ursic and Thames* [1960], were selected for study because of their similarity in soils and size. Their areas are 2.62 and 2.43 acres. Providence silt loam covers 64 per cent of field II and 25 per cent of field III. Ruston sandy loam predominates on the remainder of both units.

Measurements and records. A tier of fiberglass soil-moisture units was installed in the major soil type on each watershed. Eight depths down to 5 feet were represented. The tier on

field II was on an upper slope on Providence silt loam, that on field III occupied a midslope position on Ruston sandy loam. Readings were made at least once a week (and also after each storm) from January 1957 to March 1959 on field II and from April 1958 to March 1959 on field III.

Field-calibration samples were taken at 20 to 25 different times from a 1/20-acre circular plot surrounding each tier. At times of sampling, soil cores whose midpoint corresponded to the unit depths were taken at three points selected at random bearings and distances from the central tier.

In addition to the fiberglass unit installations, offset gravimetric plots were located on the Ruston soil of field II and on the Providence soil of field III. The purpose was to reduce the number of moisture-unit installations and to secure data from an additional location on each soil type. These plots were sampled at the same time and at the same depths as the moisture-unit plots. To check the areal representativeness of the moisture-unit and offset plots, soil samples corresponding to the unit depths were taken at intersections of a 50- by 100-foot grid over the watersheds. Such samples were secured twice during the study period.

Rainfall was measured with at least three gages on each watershed. Runoff was measured with 3-foot H-type flumes equipped with water-level recorders.

Methods. Resistance readings at the moisture-unit tiers were used to index changes in the soil-moisture status over the watershed areas. Resistance of units were related to mois-

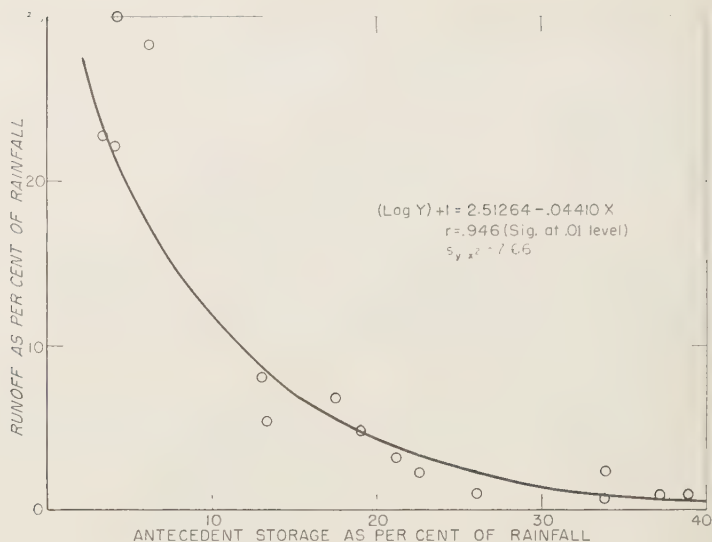


Fig. 1. Relation of runoff to pre-storm storage capacity of 0- to 6-inch soil layer.

ture contents of soil samples from the moisture-unit plots and the offset gravimetric plots [Thames, 1959]. Records of moisture content for both plots on each watershed were secured from the resistance determinations.

Summer and winter depletion curves were determined for the 0- to 6-inch layer of each type of soil [Carlson and others, 1956]. Moisture depletion from the most recent soil-moisture measurement to the day before each storm was estimated from these curves. Antecedent storage was computed as the difference between the moisture content on the day preceding the storm and the maximum moisture content recorded for the soil type. Average storage over each watershed was weighted by proportion of area in the two major soil groups.

Bulk density was determined at the same depths as soil moisture. Moisture content was converted to inches of water per 6 inches of soil, and readings of the fiberglass units were converted to the logarithm of the resistance in ohms at 60°F.

An expression of the relationship of antecedent storage and rainfall to runoff was found to be logarithmic and amenable (after transformation) to resolution into an equation with linear coefficients. Figure 1 shows the regression

developed from the records for field II. This relationship and the procedure of estimating antecedent storage for a watershed area were tested with the records for field III.

Results and discussion. The offset method of calibration was tested by calculating for each day of sampling the deviation of the gravimetric samples from the values read on the calibration curves. For each type of soil the deviations from the standard curves were compared with the deviations from the offset curves. Average deviations ranged from 0.04 to 0.06 inch of water per 3 inches of soil; 't' tests of the pooled variances indicated that differences in the deviations were not significant.

Results of the grid sampling are given in Table 1. With one exception, moisture content read from the calibration curves fell within the limits of one standard error for each soil type and soil depth on both sampling dates.

The relationship of antecedent storage and rainfall to runoff (Fig. 1) for field II shows that, under conditions of low storage opportunity, surface runoff will account for a high proportion of the rainfall. Zero storage is seldom obtained. The maximum runoff that occurred with minimum storage opportunity during the period was about 30 per cent of the total

TABLE 1. Moisture Contents of Upper 6-inch Layer as Determined from Grid Samplings and from Calibration Curves

Dates of Samples	Providence Silt Loam			Ruston Sandy Loam		
	Grid Samples			Grid Samples		
	Curve Value, in.	Mean, in.	Standard Error of Estimate, in.	Curve Value, in.	Mean, in.	Standard Error of Estimate, in.
April 24, 1959						
Offset plot	2.17	2.14	0.11	1.26	1.31	0.05
Moisture-unit plot	2.00	1.94	.08	1.31	1.33	.05
March 24, 1959						
Offset plot	1.81	1.72	.13	1.10	1.15	.04
Moisture-unit plot	1.92	1.95	.15	1.08	1.06	.07

m rainfall. The Ruston-Providence soil complex has high transmission rates, and the bulk of most rains percolated to depths below 6 inches.

For dry conditions in the Ruston-Providence complex, Figure 1 indicates that, as storage opportunity increases to about 40 per cent of storm amount, runoff will be negligible.

Even though storage space is considerable, it may remain unused because water cannot move through it rapidly enough. Under this condition rainfall intensity becomes a factor in determining the usability of storage and the volume of runoff. Much of the variation from the regression is probably due to variations in rainfall intensity. For instance, a 2.10-inch storm on August 8 (omitted from the analysis) had a maximum 15-minute intensity of 3.40 inches per hour as compared with an average of 1.52 inches for all other storms; 95 per cent of the rain fell within one hour. Although the available storage in the 0- to 6-inch soil layer on that day before the storm exceeded the total rainfall, 8.8 per cent of the precipitation ran

Runoff prediction. The validity of using the antecedent moisture in the 0- to 6-inch layer to improve runoff predictions is indicated by the comparisons in Table 2. Runoff was predicted for field III with the relationship developed for field II and with the antecedent storage and rainfall measurements made on field III. Greatest accuracy was obtained with the latter storms. Average deviation between pre-

dicted and actual runoff for the storms of November through February was 0.042 inch, as compared with a mean of 0.082 inch for all storms. The greatest deviation occurred with a 5.75-inch storm that began September 19 and extended through September 21. Although the greatest errors occurred with summer storms, the differences between predicted and actual runoff were not significant (by 't' test) either within seasons or for the entire period.

Conclusions. Amount of rain and available storage in the soil before the rain were two major influences on the amount of surface runoff. The amount of water held in the 0- to

TABLE 2. Comparison of Actual and Predicted Runoff, Field III

Date	Rainfall, in.	Antecedent Storage, in.	Actual Runoff, in.	Predicted Runoff, in.
May 10	0.84	0.22	0.02	0.02
June 1	1.46	.25	.04	.08
Aug. 8	2.10	.42	.18	.09
Sept. 2	1.71	.24	.05	.14
Sept. 19	5.75	.12	2.04	1.60
Nov. 14	1.76	.34	.03	.08
Jan. 14	2.09	.15	.44	.33
Jan. 20	1.00	.10	.12	.12
Feb. 8	.77	.19	.00	.02
Feb. 11	1.06	.15	.03	.08
Feb. 13	1.71	.05	.44	.42
Total	20.25		3.39	2.98

6-inch soil layer provided a practical index to the antecedent-storage condition of the soil profile. The study suggests that reasonably accurate predictions of storm runoff can be computed from rainfall data if reliable estimates of the available storage in the top 6 inches of soil are made.

Runoff predictions using the 0- to 6-inch soil storage index were most accurate for winter periods and least accurate for high-intensity storms.

REFERENCES

- Carlson, C. A., K. G. Reinhart, and J. S. Horton, Predicting moisture in the surface foot of soil, *Proc. Soil Sci. Soc. Am.*, 20, 412-415, 1956.
- Hartman, M. A., R. W. Baird, and J. B. Pope, retention index for rainfall runoff relationships, *U. S. Dept. Agr., Agr. Research Serv., Research Rept.*, 309, 10 pp., 1958.
- Thames, J. L., Measuring soil moisture over large areas with single installations of moisture units, *J. Geophys. Research*, 64, 257-262, 1959.
- Ursic, S. J., and J. L. Thames, Effect of cover types and soils on runoff in northern Mississippi, *J. Geophys. Research*, 65, 663-667, 1960.
- Whelan, D. C., L. E. Miller, and J. B. Cavalleri, A method of determining surface runoff by routing infiltrated water through the soil profiles, *U. S. Forest Serv., Northeast. Forest Experiment Sta., Sta. Paper 54*, 15 pp., 1952.
- (Manuscript received August 6, 1959; revised November 25, 1959; presented at the Fortieth Annual Meeting of the American Geophysical Union, May 1959, Washington, D. C.)

Translocation of Moisture with Time in Unsaturated Soil Profiles

PAUL R. NIXON AND G. PAUL LAWLESS

*Agricultural Research Service, U. S. Department of Agriculture
Lompoc, California*

Abstract. Downward translocation of moisture in soil profiles under various types of natural vegetation and a denuded plot was observed during a prolonged rainless period. Moisture determinations were made to 20-foot depths with a neutron-scattering moisture meter. The observations were made as part of a study of ground-water recharge by deep penetration of rain water. The significant magnitude that translocated moisture may reach is illustrated by data obtained in sand under brush cover. In this case, deeply translocated moisture was equal to 159 per cent of evapotranspiration during the first rainless month. Approximately 31 per cent of the moisture content of a 20-foot profile under a denuded plot was lost by downward movement from the first to the 240th day of a rainless period. The moisture content W of various soil layers under the plot varied with time T according to the relation $W = a T^{-b}$.

Introduction. Translocation of moisture in soil profiles was observed in connection with study of ground-water recharge by rain. The study area was located 130 miles west-northwest of Los Angeles, California, in a coastal climate. This area has a short rainy season, which coincides with the winter months, followed by a long dry season. All moisture data reported in this paper were obtained with a neutron-scattering moisture meter from a single access tube at each site. These neutron data were corrected for the air-soil interface effect.

Soil-moisture changes under brush. Accretion of soil moisture due to the penetration of rain water into sand under a brush cover of chamise is shown in Figure 1. Dates of sampling and information on precipitation and soil moisture related to Figures 1 through 3 are shown in Tables 1 and 2. The initial penetration of rain water advanced as a wetting front. By the end of the rainy season, moisture had penetrated to a depth of 13 feet. The total precipitation during the rainy season was 24 inches. Average precipitation for the area is 13.5 inches. The year 1958, with 24 inches of rainfall, was the wettest since 1941. It is probable, barring any past changes in vegetation, that during the previous 16 years rain water had never penetrated so deeply (Figs. 1 through 3). If there were roots below the sampled depth, they very

likely had depleted the moisture to the depth of rooting before this study was begun.

Translocation of moisture following the rainy season is illustrated by Figure 2. During this

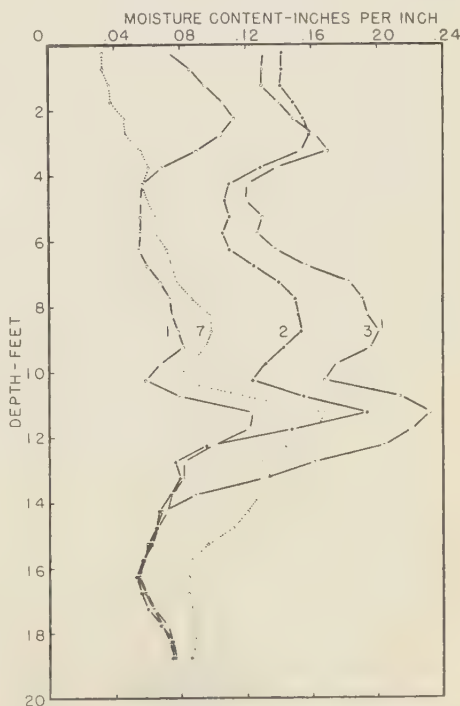


Fig. 1. Moisture accretion in Marina sand under brush.

Contribution from Soil and Water Conservation Research Division, Agricultural Research Service, U. S. Department of Agriculture.

TABLE 1. Antecedent Precipitation at Brush-Covered Site

Sampling Date	Curve Identification	Antecedent Precipitation, inches			
		1 day	2 days	5 days	10 days
1957					
Dec. 11	1	0.02	1.40	1.40	1.40
1958					
Feb. 25	2	1.12	1.12	1.14	3.37
April 8	3	0	0.88	2.20	5.24
May 9	4	0	0	0	0
June 18	5	0	0	0	0.04
Aug. 26	6	0	0	0	0
Nov. 18	7	0	0.01	0.13	0.16

TABLE 2. Soil Moisture Changes and Precipitation at Brush-Covered Site

Sampling Period	Period Identification	Days in Period	Precipitation, inches	Soil-Moisture Change, inches
Dec. 11-Feb. 25	1-2	76	9.98	+8.34
Feb. 25-Apr. 8	2-3	42	9.99	+5.12
Apr. 8-May 9	3-4	31	0.02	-2.18
May 9-June 18	4-5	40	0.31	-2.63
June 18-Aug. 26	5-6	69	0	-3.92
Aug. 26-Nov. 18	6-7	84	1.74	-2.17

dry season, downward penetration of moisture continued without appreciable addition of rain water to the profile. Water advanced 3 feet from the first to the thirty-second day following the last rain of the wet season (see curves 3 and 4). During the next 40 days (curve 4 to curve 5) an additional $1\frac{1}{2}$ feet of penetration occurred, with successively decreasing rates of advance thereafter.

The changes in moisture content, for each depth, between the profiles of Figure 2 are shown in Figure 3. Curve 3-4 for the first 31-day period of the dry season shows that more moisture was translocated downward than was lost by evapotranspiration. The loss of moisture from the upper 11.5 feet of the profile amounted to 5.68 inches. Downward translocation caused a gain in moisture below that depth amounting to 3.50 inches. Calculated evapotranspiration of 2.20 inches was determined from the sum of the net loss of soil moisture of 2.18 inches ($5.68 - 3.50$) and the precipitation of 0.02 inches which occurred during period 3-4. Thus, during the first month of the dry season the

quantity of deep translocated moisture was 15 per cent of the amount lost by evapotranspiration (3.50 inches versus 2.20 inches).

Transient moisture. Soil-moisture changes during additional dry-season periods are shown by the curves 4-5, 5-6, and 6-7 in Figure 3. Deep translocation continued throughout the dry season, but at a decreasing rate. During period 4-5 the net moisture change in the 12- to 16-foot zone was a slight loss, although large gains had been made in this zone during the previous period. The gain in moisture in the 16- to 18-foot zone was an indication of the continuation of the deep translocation with time.

From the curves of Figure 3 it is apparent that moisture extraction, which was initially confined to the upper portion of the profile, moved downward with time until the last period illustrated, when the major extraction of water by brush occurred in the 10- to 16-foot zone.

In instances when soil-moisture measurements were well timed with respect to rainfall, it was

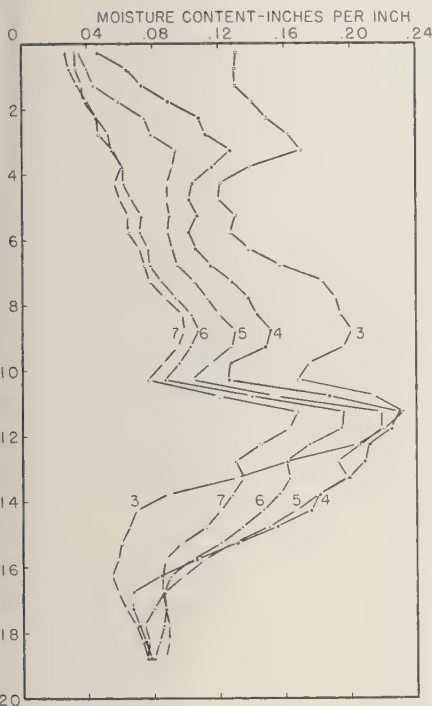


Fig. 2. Moisture depletion in Marina sand under brush.

observed that the passage of moisture through recently wetted profiles caused but a relatively slight increase in moisture content at each depth. The increase of moisture in the soil profile under a denuded plot is shown in Figure 4. The curve for February 17 represents soil moisture in a rather dry period within the

rainy season. The curve for February 27 shows the effects of moderately wet antecedent conditions. Antecedent precipitation is shown in Table 3.

Translocation of moisture under a denuded plot. Dry-season data, obtained at the denuded plot mentioned above, are represented by Figures 5 and 6 and Table 4. Figures 5 and 6 were prepared from the same data. The information on moisture of October 16 is omitted from Figure 5 for the sake of clarity.

In Figure 6 soil-moisture contents are plotted against time on logarithmic scales. The straight-line relations indicate that the rate of water loss for the various soil layers was proportional to water content and inversely proportional to the time since the end of the rainy season [Richards and Richards, 1957].

This relationship of moisture content and time, demonstrated in Figure 6, is expressed by the function

$$W = aT^{-b}$$

where W is the moisture content of the soil layer, expressed as the equivalent depth of water in inches, and T is time in days since wetting. The constants a and b are peculiar to the soil layer.

Moisture content versus time. Data shown in Table 5 are the constants of the exponential function (above) for various soil layers of the denuded plot. The a constant is the moisture content of the layer, in inches, 1 day after the rains. Values of the a constant were determined graphically from Figure 6 by extrapolating the lines, fitted by eye, back to the first day of the dry season ($T = 1$).

TABLE 3. Antecedent Precipitation at Denuded Plot

Sampling Date	Curve Identification	Antecedent Precipitation, inches			
		1 day	2 days	5 days	10 days
1958					
Feb. 17	...	0.01	0.02	0.02	0.58
Feb. 27	1	0.10	1.22	1.22	3.46
Apr. 16	2	0	0	0	1.96
May 13	3	0	0.01	0.01	0.01
July 9	4	0	0	0	0
Oct. 16	5	0	0.01	0.01	0.02
Dec. 3	6	0	0	0	0

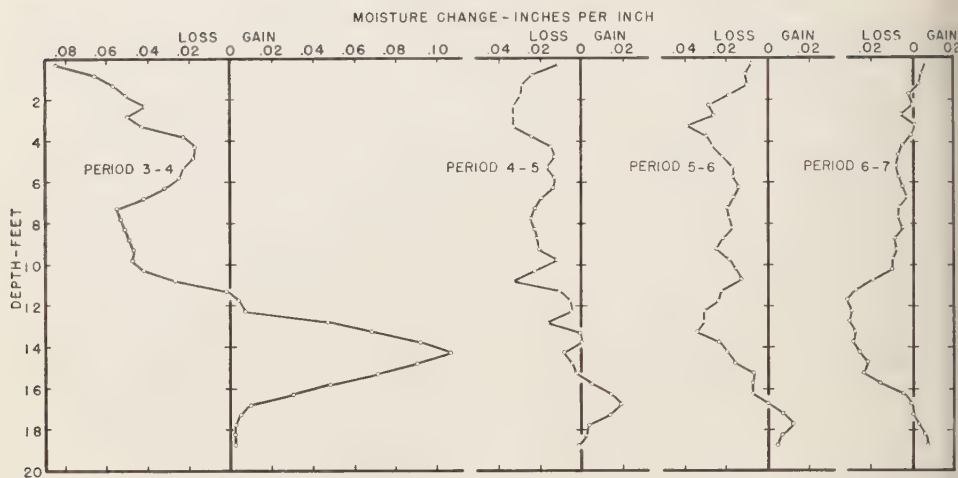


Fig. 3. Moisture changes in Marina sand under brush; dry season, 1958.

The b constant is related to the rate at which moisture was lost from the soil. The increase in slope of the lines in Figure 6, representing successively deeper profiles, is indicated by the increase of the b constant in the left-hand portion of Table 5. However, if successive 2-foot layers are studied separately it is seen that, under the conditions prevailing at the denuded plot, the b constants increased to the 10- to 12-foot layer and decreased below that level (see right-hand portion of Table 5). This apparent discrepancy between the two sets of numbers can be understood when it is realized that the moisture content under the denuded plot increased with depth.

Evaporation versus drainage. Two factors contributed to the loss of moisture: evaporation from the surface and deep translocation or drainage out of the measured profile. From the

data collected it is impossible to make a precise separation between evaporation and deep translocation. Surface evaporation noticeably reduced the moisture content of the top 2 feet. The rate of moisture reduction of this surface layer was almost constant rather than being a negative exponential function of time. This is illustrated by the dashed line in Figure 6.

A study of moisture content with respect to time, by $\frac{1}{2}$ -foot depth increments, revealed that the slope of the lines when plotted on log-log paper tended to decrease down to the 4-foot depth. This indicates that proportionally less and less water was lost with increasing distance to 4 feet below the surface. From the 4- to 12-foot depths, the b constants of the 2-foot layers increased (right-hand column of Table 5), indicating a greater loss of moisture in proportion to initial content, in successively lower

TABLE 4. Soil-Moisture Changes and Precipitation at Denuded Plot

Sampling Period	Period Identification	Days in Period	Precipitation, inches	Soil-Moisture Change, inches
Feb. 27-Apr. 16	1-2	48	9.99	-2.27
Apr. 16-May 13	2-3	27	0.07	-3.42
May 13-July 9	3-4	57	0.16	-2.54
July 9-Oct. 16	4-5	99	1.45	-0.96
Oct. 16-Dec. 3	5-6	48	0.27	-0.75

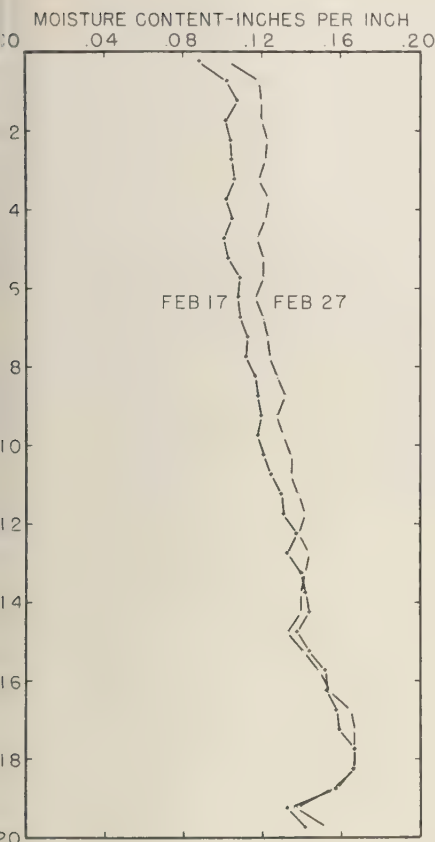


Fig. 4. Effect of antecedent precipitation on soil moisture in Marina sand under a denuded plot.

ers. Below the 12-foot level the moisture es, by layers, became proportionally smaller. Tensiometer data were collected only at pths of 1, 2, and 3 feet. Maximum suctions served for these depths were 35, 20, and 17 atibars with respect to the ground surface. he small upward potential between the 3- and 2-foot levels supports the belief that only all quantities of water were conducted up- rd from below 3 feet in the sand, with little tribution to evaporation from below 4 feet. If one assumes that all moisture lost from e upper 4 feet of the profile was lost by evap- tion and all moisture lost from below that pth went to deep translocation, values shown Table 6 indicate that 11.2 inches passed

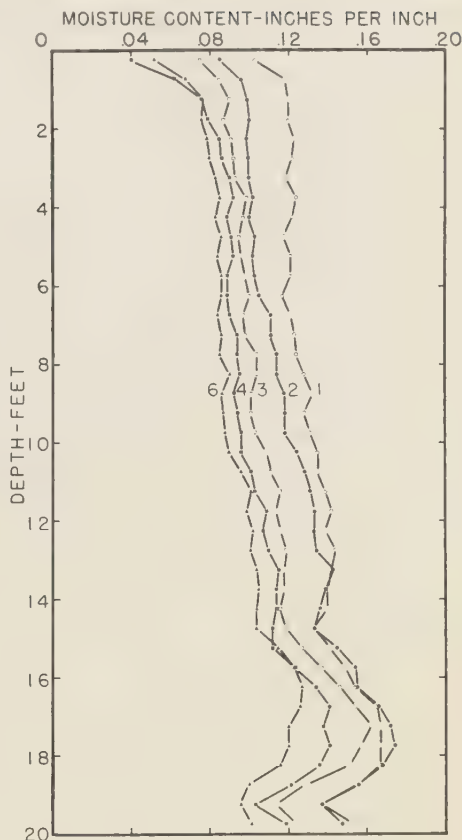


Fig. 5. Moisture depletion in Marina sand under a denuded plot.

downward out of the 20-foot profile from the first to the 240 day of the dry season. The right-hand column of Table 6 shows the amount of water that passed through the soil during the dry season; the quantity increased with depth. Thus 9.6 inches of water entered the 18- to 20-foot layer from above. This amount plus 1.6 inches contributed by the layer itself resulted in a total downward loss from the profile of 11.2 inches of moisture.

In the table, the 1-day moisture contents are theoretical, being the a constants of the soil layers of Table 5. The 240-day moisture contents are those which were observed in the field. On the basis of these assumptions, 31 per cent of the first-dry-day moisture content of

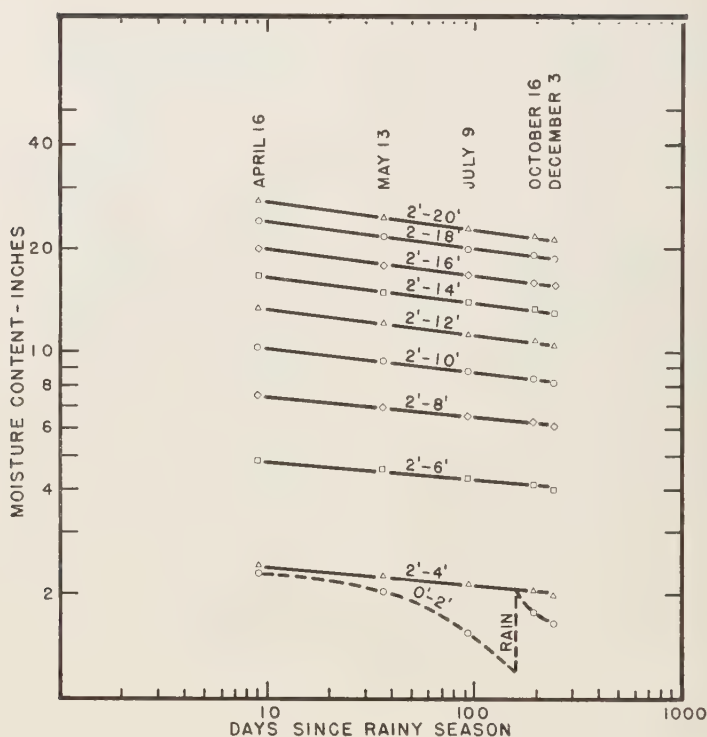


Fig. 6. Depth of water in various layers of Marina sand under a denuded plot.

35.6 inches drained from the profile during the 239-day period.

Summary. The data here presented, obtained from nonirrigated soil, illustrate the significant magnitude which translocated soil moisture assumes under certain field conditions.

TABLE 5. Constants of the Exponential Function Relating Moisture Content to Time

Soil Profile, ft.	<i>a</i>	<i>b</i>	Soil Layer, ft.	<i>a</i>	<i>b</i>
2-4	2.7	0.054	2-4	2.7	0.054
2-6	5.4	.054	4-6	2.7	.054
2-8	8.6	.062	6-8	3.2	.078
2-10	12.0	.068	8-10	3.4	.085
2-12	15.9	.076	10-12	3.9	.106
2-14	19.9	.080	12-14	4.0	.094
2-16	24.2	.081	14-16	4.3	.086
2-18	28.9	.081	16-18	4.7	.084
2-20	33.0	.081	18-20	4.1	.083

These observations are in general agreement with measurements made in irrigated soils by *Robins and others* [1954] and experimental studies by *Richards and others* [1956].

Translocation of moisture has been observed at 13 sites under study by the authors. Under natural conditions, the quantity of translocated moisture has been greatest in coarse-textured soil. Nevertheless, in soils ranging in texture from sand to clay loam measurable migration was observed.

Reference in this paper has been made primarily to downward translocation of moisture. Movement would, of course, always be in the direction of the potential gradient.

Conclusions. Translocation may frequently assume proportions which cannot be neglected in hydrologic studies involving soil moisture. Under favorable conditions, moisture translocated downward in unsaturated soil may contribute to ground-water recharge.

TABLE 6. Loss of Soil Moisture by Deep Translocation during the period of 1 to 240 Days Following the Rainy Season

Soil Layer, ft	1-Day Moisture, in.	240-Day Moisture, in.	Loss of Moisture, in.	Cumulative Loss, in.
2-4	2.7	2.0	0.7	...
4-6	2.7	2.0	.7	0.7
6-8	3.2	2.0	1.2	1.9
8-10	3.4	2.1	1.3	3.2
10-12	3.9	2.3	1.6	4.8
12-14	4.0	2.5	1.5	6.3
14-16	4.3	2.7	1.6	7.9
16-18	4.7	3.0	1.7	9.6
18-20	4.1	2.5	1.6	11.2

Acknowledgments. This study was made in cooperation with the U. S. Geological Survey, U. S. Bureau of Reclamation, Santa Barbara County California Water Agency, California State Department of Water Resources, and the California Agricultural Experiment Station.

REFERENCES

- Richards, L. A., W. R. Gardner, and G. Ogata, Physical processes determining water loss from soil, *Soil Sci. Soc. Am. Proc.*, 20, 310-314, 1956.
- Richards, L. A., and S. J. Richards, Soil moisture, *U. S. Department of Agriculture Yearbook*, 49-60, 1957.
- Robins, J. S., W. O. Pruitt, and W. H. Gardner, Unsaturated flow of water in field soils and its effect on soil moisture investigations, *Soil Sci. Soc. Am. Proc.*, 18, 344-347, 1954.
- (Manuscript received August 22, 1959; revised November 24, 1959; presented at the Fortieth Annual Meeting of the American Geophysical Union, May 1959, Washington, D. C.)

Effect of Cover Types and Soils on Runoff in Northern Mississippi

S. J. URSIC AND J. L. THAMES

*Southern Forest Experiment Station
Forest Service, U. S. Department of Agriculture
Oxford, Mississippi*

Abstract. Hydrologic and meteorologic data for individual storms were collected from small headwater catchments representing three types of cover in northern Mississippi during 1958. Surface runoff and peak flows were greatest from abandoned fields, intermediate from depleted upland hardwood forests, and least from 20-year-old loblolly pine plantations that had been established on eroding farm land. The pine cover has been a highly effective flood-abatement measure.

The soils are silt loams of loessial origin and Coastal Plain sandy loams. Surface runoff from each of the three types of cover increased directly with the proportion of loessial soil. A shallow fragipan more than doubled the amount of surface runoff and increased peak flows. Detention storage above the fragipan was directly related to flow stage at the flume.

Introduction. The effect of several types of forest and potential forest cover on surface runoff and erosion is being studied in the vicinity of Oxford, Mississippi, by the Oxford Research Center of the Southern Forest Experiment Station. The experimental watersheds are abandoned fields, upland hardwood forests, and pine plantations. The research is in cooperation with the Yazoo-Little Tallahatchie Flood Prevention Project, one of the largest erosion- and flood-control programs in the eastern United States. This paper is limited to data collected during 1958.

Although the beneficial effects of an undisturbed forest cover on watersheds are well established, there is a dearth of information concerning water yield under various kinds of vegetation. The effect of the underlying soils on small-watershed hydrology has been almost completely neglected. In northern Mississippi the undisturbed forest remains. The area has been heavily row-cropped, and a large percentage of the highly erodible soils have been abandoned and are reverting to forest.

Climate. The annual precipitation at Oxford averages 52 inches and is fairly well distributed throughout the year. Measurable precipitation occurs approximately one day in four. The growing season lasts 218 days. Temperatures average 45°F in January and 80°F in July. Occa-

sionally the soils freeze to shallow depths. Brief convective summer storms of high intensity are common. Most of the winter precipitation is cyclonic.

Vegetative cover. The three types of cover, each replicated three times, represent a large total acreage of the upland portion of northern Mississippi and environs.

The abandoned-field watersheds, once cotton land, support a grass-herbaceous cover dominated by broomsedge (*Andropogon* spp.). As determined by line transects, 10 per cent of the ground surface is covered by grasses and forbs and 30 per cent by litter; the remaining 60 per cent is exposed soil.

The upland hardwood forest watersheds, depleted by decades of grazing, wildfire, and overcutting, support post oak (*Quercus stellata* Wangenh.), blackjack oak (*Quercus marilandica* Muenchh.), and hickory (*Carya* spp.), all of low commercial value. The understory is sparse, and reproduction of desirable trees is lacking. Basal area (sum of cross-sectional areas of tree stems determined from diameters measured at 4.5 feet above the ground) in trees over 3.55 inches in diameter averages 48 square feet per acre, about half of which is cull.

The 20-year-old stands of loblolly pine (*Pinus taeda* L.) average 77 square feet of basal area and 1000 cubic feet of volume per

TABLE 1. Total Annual Runoff, 1958

Types of Cover	Precipitation Average, in.	Runoff from Watersheds		Average Runoff, in.
		Loess Soils, in.	Loess and Coastal Plain Soils*, in.	
Abandoned fields	55.98	14.86	4.80	8.16
Hardwoods	56.50	7.50	4.68	5.62
Pine	52.94	3.17	0.30	1.26

* Average of two watersheds.

acre in trees over 3.55 inches in diameter. The ground surface is covered with a heavy mat of pine litter. The understory averages 320 trees per acre. It is dominated by sweetgum (*Liquidambar styraciflua* L.) and eastern redcedar (*Juniperus virginiana* L.). Before the pines were planted, the watersheds were primary silt sources. It is estimated that as much as 2 feet of the surface soil had been removed by erosion. Several gullies are 5 feet below the present soil surface.

All the watersheds are protected against grazing and have not been burned over for at least 7 years.

Soils. Two major groups of red-yellow podzolic soil are represented: loessial soils by the Loring, Providence, and Lexington series, and upper Coastal Plain soils primarily by the Ruston series. Similar loessial soils are prevalent in all States bordering the Mississippi River from Illinois southward. The Ruston soils, together with closely allied series, are important agricultural and forest soils in all the Atlantic and Gulf Coastal States from Virginia to Texas.

The loessial soils on the watersheds have a silt loam surface and the Coastal Plain soils are sandy loams. The Providence soil in the uneroded profile is characterized by a fragipan at the 18- to 24-inch depth. The other soils have moderate to good internal drainage.

Abandoned field I, Hardwood III, and Pine III watersheds are completely loess; the first two are Providence silt loams, and Pine III is Loring-Lexington silt loam. The remaining watersheds have a combination of soils, with the loess on the ridges and upper slopes grading into the sandy loam textures on the middle slopes. Loessial soil on these units varies from 29 to 65 per cent of the watershed area.

The maximum difference in elevation from the head of the watersheds to the flume site ranges from 45 to 70 feet on the mixed-soil watersheds and from 35 to 45 feet on the more gently sloping loessial units.

Instrumentation and records. Surface runoff¹ is measured with 3-foot H-type flumes equipped with water-stage recorders. Deposited sediment is collected in a concrete approach section. Samples for suspended-sediment determinations are obtained with a wheel sampler. Precipitation is measured with a network of recording and standard gages—about one gage per acre. Soil moisture is recorded on each major soil-cover complex with fiberglass moisture units supplemented by gravimetric sampling. Relative humidity and temperature are measured in standard U. S. Weather Bureau shelters.

As far as possible, each watershed is serviced after every rain. Storms are separated on the basis of a 6-hour lapse between successive precipitations.

The abandoned-field and depleted-hardwood forest watersheds are within a quarter-mile radius. The pine plantation watersheds are similarly arranged but are some 7 air miles distant and are sometimes subject to a differing rainfall regime, especially for spring and summer convective storms. This variation in rainfall pattern was carefully considered and found

¹ The part of runoff that passes over the surface is called surface runoff or overland flow. Surface runoff, in distinction from flow in channels, persists only a short time after rain stops and includes primarily sheet flow and that in incipient surface channels. Where a fragipan is present, runoff also includes considerable subsurface quick return flow.

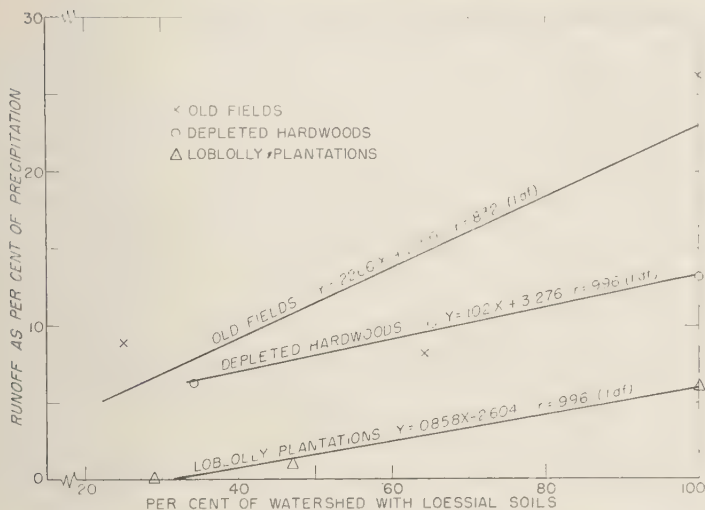


Fig. 1. Surface runoff as influenced by soils.

unimportant in the discussions and conclusions offered in this article.

Results. Total annual surface runoff in area is given in Table 1. Precipitation for the year approximated the long-term average of 52 inches, but distribution was atypical, with over 10 inches in both April and September—two times three times normal.

To determine the effect of the two major soil types, runoff as a percentage of precipitation was plotted over the percentage of watershed area with loessial soils and regressions fitted. Figure 1 indicates that the amount of runoff is greater from the finer-textured loessial soils for all three types of cover. Although none of the regressions are significant (because of inadequate sampling—only 3 observations per regression), the correlation coefficients for the depleted hardwood and pine cover types ($r = 0.996$) closely approximate the coefficient ($r = 0.997$) required for the 5 per cent level of significance.

Figure 1 also shows that runoff decreases in the order: abandoned fields, depleted upland hardwood forests, and loblolly pine plantations. The deeper and heavier ground litter on the loblolly plantation watersheds curtailed runoff, probably by protecting and maintaining conditions favorable for absorption at the soil sur-

face. In an earlier study in this area, *Meginnis* [1934] measured runoff from an abandoned-field plot that had been covered with the floor of a mature oak forest; the treatment reduced runoff by more than half during the first year. *Kittredge* [1954], in a 14-year study, found that surface runoff as a percentage of annual seasonal precipitation ranged between 0.2 and 9.6 from pine plots and between 0.01 and 36.4 from plots with annual grasses. *Rowe* [1955] reported that as little as 0.5 inch of pine litter was effective in controlling surface runoff.

Of special interest is the considerable surface runoff from the abandoned-field watershed with the Providence fragipan soil. Long periods of post-storm flow prompted the installation of a series of shallow wells at eight locations approximately 30 feet apart and at 4-foot contour intervals extending from near the flume to the ridge. Three 3-inch perforated gutter pipes were installed at each location, to depths of 3 inches, 6 inches, and several inches below the top of the fragipan (which was 16 to 25 inches below the surface). Depth of water in the wells was measured to $\frac{1}{2}$ inch after each storm.

The wells at the 3- and 6-inch depths seldom contained free water, but those extending into the fragipan gave definite indication of a

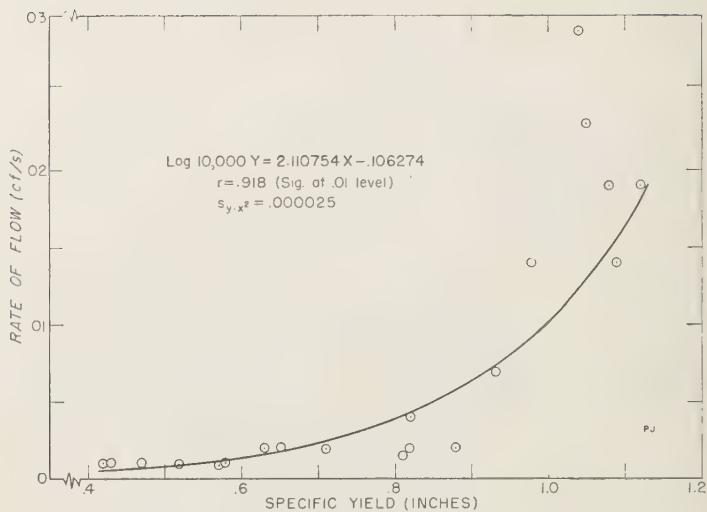


Fig. 2. Relation between rate of surface flow and specific yield.

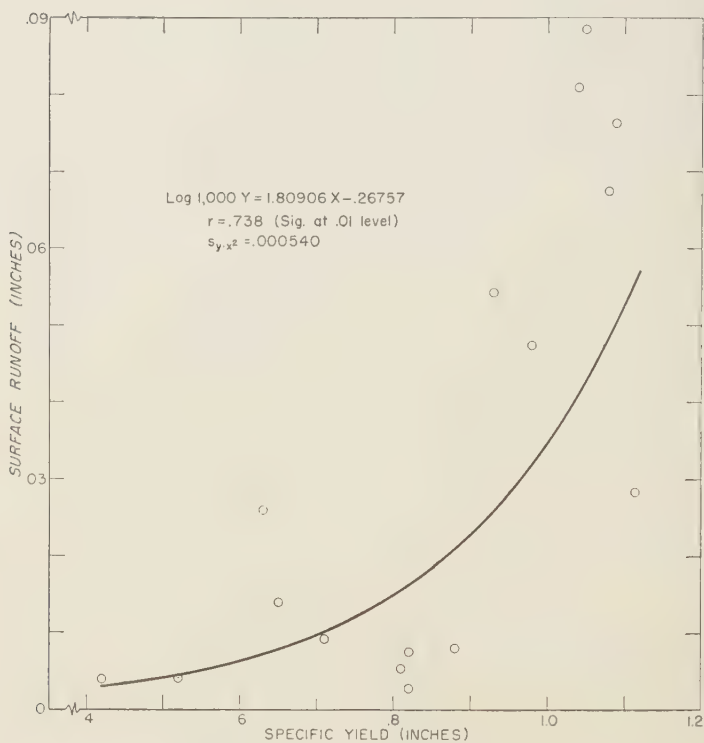


Fig. 3. Relation between specific yield and surface runoff.

hed water table. Yield at the flume site was inches greater from this watershed than from other abandoned-field watersheds, which have mixed soils (Table 1).

To explore the effect of the fragipan on surface-water yields, the average depth of the detention water was converted to specific yield, and the detention storage capacity or that of soil water available to stream flow, and related to the rate of flow in ft³/sec (cubic feet per second) at the flume.

The relationship derived from rectified log plots for Y is shown in Figure 2. The basis is measurements from February through July 1958. It is probable that rates of flow greater than 0.020 ft³/sec have a surface runoff component, as they were read from the more rapid receding limbs of the hydrographs.

To determine how much of the soil water above the pan would appear as surface runoff, the volumes of runoff occurring after the rains were measured were plotted over specific yield. This relation for 16 storms during February to July 1958 is shown in Figure 3. Here the difficulties of estimating the cessation of runoff at low flows, the seasonal differences, the variable detention storage in the soil above the pan level immediately after large storms, and the sometimes discontinuous data from subsequent storms can all contribute to possible error in measurement. Nevertheless, the highly significant correlation coefficients indicate (if it can be assumed that successive observations on the same unit are independent—a fundamental requirement of least-squares analyses) that both runoff rate (Fig. 2) and volume (Fig. 3) of post-storm flow are closely related to the amounts of detention water held over the fragipan.

For the 16 post-storm periods, a maximum of 10 per cent of the detention water available for stream flow appeared at the flume site. This would indicate that the fragipan on the Providence soil restricts but does not prevent percolation of water. The conclusion is supported by field observations, for the soil is wetted to a depth of at least 10 feet during the winter recharge period. On a morphologically comparable soil in Missouri, Fletcher and McDermott

[1957] found that the fragipan restricted water transmission and also limited the depth of root penetration. Carlisle and others [1957] found that fragipan soils have wide geographic distribution in the United States.

The influence of the fragipan on total surface-water yield is also evident on the hardwood-forest watershed with Providence soil (Table 1). However, the deeper, less-eroded soils with a greater storage opportunity above the pan presumably reduce the effect of the pan.

Peak flows within each type of cover generally vary directly with the amounts of storm runoff. When the abandoned-field and hardwood-forest watersheds are paired on the basis of similar soils, the peaks from the forest sites are always less than half the peak discharges from the abandoned fields, even though the amounts of runoff may be similar. Flows from the pine plantation watersheds occur only from very intense rains, and the peaks are lower than from the other types of cover.

Production of sediment was low from the three types of cover—all less than 1/2 ton per acre per year. Loss of soil was highest from the abandoned fields and insignificant from the pine plantations.

REFERENCES

- Carlisle, F. J., E. G. Knox, and R. B. Grossman, Fragipan horizons in New York soils: I. General characteristics and distributions, *Proc. Soil Sci. Soc. Am.*, **21**, 320-321, 1957.
- Fletcher, P. W., and R. E. McDermott, Moisture depletion by forest cover on a seasonally saturated Ozark ridge soil, *Proc. Soil Sci. Soc. Am.*, **21**, 547-550, 1957.
- Kittredge, J., Influences of pine and grass on surface runoff and erosion, *J. Soil and Water Conserv.*, **9**, 179-186, 1954.
- Meginnis, H. G., Influence of forest litter on surface runoff and soil erosion, *Rept. Am. Soil Survey Assoc. (Rept. Ann. Meeting 15)*, **18**, 115-118, 1934.
- Rowe, P. B., Effects of the forest floor on disposition of rainfall in pine stands, *J. Forestry*, **53**, 342-348, 1955.

(Manuscript received August 6, 1959; revised November 25, 1959; presented at the Fortieth Annual Meeting of the American Geophysical Union, May 1959, Washington, D. C.)

Radioisotopes P^{32} , Be^7 , and S^{35} in the AtmosphereD. LAL,¹ RAMA, AND P. K. ZUTSHI*Tata Institute of Fundamental Research, Bombay, India*

Abstract. We have made simultaneous determinations of the concentrations of the radioisotopes P^{32} , Be^7 , and S^{35} in several rains collected at Bombay during the monsoon period of 1958.

The observed annual deposition rates of P^{32} and Be^7 are found to agree within the expected meteorological fluctuations with those measured in the preceding two years at Bombay and other stations in India. S^{35} deposition, however, has varied significantly outside normal fluctuations. From the observed variations in the annual fallout of these isotopes during 1956-1958 and relative isotope concentrations in individual rains, it has been concluded that nuclear weapons have not resulted in any appreciable contributions to the observed Be^7 and P^{32} activities. However, S^{35} was contributed in appreciable quantities. During 1958, S^{35} concentrations were very large, amounting to as much as 35 per cent of the observed Sr^{90} concentrations in some rains.

The relative concentrations of the isotopes Be^7 and P^{32} in rains vary in a manner that would be expected for their production in the troposphere by cosmic rays, and their removal by wet precipitation with an average removal period of about 40 days. None of the rain samples analyzed so far could have resulted in precipitation occurring from an air mass which descended in the troposphere after an irradiation in the stratosphere for time periods that are long compared with the half-lives of the isotopes.

Introduction. Several cosmic-ray-produced isotopes of half-lives appropriate for use as tracers for studying the general atmospheric circulation processes have recently been detected in rain water. Isotope production rates in different regions of the atmosphere have been calculated by Lal and others [1958]. The relative production rates are found to depend not only on the altitude for a given latitude, but also appreciably on latitude for a given altitude. The relative production rates of various isotopes of interest, however, do not vary by more than 10 per cent, with position in the atmosphere. Because of these features in the cosmic-ray production of isotopes, one can hope to study the gross circulation features characterizing the removal of material from the stratosphere and the troposphere by extended observation of their deposition rate as a function of position on the earth's surface. The detailed mechanisms of exchange of air between the stratosphere and the troposphere, and scavenging by precipitation, can also be

studied by observations of the ratios of concentrations of isotopes having widely different half-lives in individual air masses [Lal and others, 1958].

Isotopes suitable for meteorological studies are P^{32} , Be^7 , S^{35} , and Na^{22} ; their half-lives are 14.3 days, 25 days, 53 days, 87 days, and 2.6 years, respectively. To date, the experimental work has chiefly been confined to studying them in wet precipitation because of the difficulties encountered in quantitative measurements of radioactivity in air. All but Na^{22} have been measured since 1956 in wet precipitation occurring at several stations in India. The results are representative of the annual precipitation in the case of Be^7 and P^{32} at Bombay and Kodaikanal [Goel and others, 1959; Rama and Zutshi, 1958]. Peters [1959] and Goel and others [1959] showed that the various measurements available until 1957 were consistent with the interpretation that most of the observed activity in rain water was derived from air irradiated in the troposphere. The observed fallout rates agreed reasonably well with those calculated by Lal and others [1958] for P^{32} , P^{33} , and Be^7 . The S^{35} fallout was, however,

¹ Present address: Scripps Institution of Oceanography, University of California, La Jolla, California.

found to be appreciably higher than expected, and it could not easily be reconciled with the calculations.

The results of Goel and his co-workers also showed that in 11 rain samples analyzed at Bombay variations in the ratios of the concentrations of Be^7 to P^{32} are confined to a much smaller range than observed for their absolute concentrations. Absolute isotope concentrations are evidently dependent upon the amount of moisture condensed from a given air mass, whereas isotope ratios are independent of the amount of moisture condensed per gram of air, as long as condensation removes the activities with high efficiency. Therefore, the various isotope ratios in rain samples will depend only upon the trajectory of the air mass under consideration and not upon local meteorological conditions. The observed distribution in the ratio $\text{Be}^7/\text{P}^{32}$ was found to be consistent with that expected for removal of these activities from air irradiated in the troposphere with an average period of 35 days; none of the samples had characteristics of air descending from the stratosphere.

We have now made simultaneous determinations of the concentrations of S^{35} , Be^7 and P^{32} in 19 rain samples collected during the monsoon period of 1958. In some of these samples we have also determined the Sr^{90} and Sr^{90} concentrations. The observed fallout of Be^7 and P^{32} is found to agree, within expected meteorological fluctuations, with that during 1956 and 1957; the fallout of S^{35} , however, was considerably higher, especially in rains precipitated after the first week of July 1958. The results showed that the observed Be^7 and P^{32} activities were still in large part of cosmic-ray origin, whereas the S^{35} activity was largely due to nuclear weapons tests.

Considering only the Be^7 and P^{32} data, we have studied the general features of the mechanism of removal of radioactivity from the atmosphere.

Experimental procedure. In order to test the reliability of the analyses, we made three independent measurements for each rain sample. Rain water samples of 170 liters were collected: pH was first adjusted to 2 to 3, and the water was then divided into three equal parts of 48 liters each and one of about 25 liters.

From each of the 48-liter samples phosphorus, beryllium, and sulfur activities were extracted and counted separately. The chemical procedure and the counting techniques employed have been described earlier [Goel and others, 1958].

Sr^{90} and Sr^{90} activities were determined in the 25-liter sample following the procedure developed by Osmond and others [1957]. Most of the samples were subjected to an additional purification step, and the Sr^{90} , Sr^{90} activities were redetermined. The two sets of results always agreed within 20 per cent. Mean values are reported in this paper.

Results. In all three samples the determinations of S^{35} , Be^7 , and P^{32} activities usually agreed within the experimental errors (~ per cent). However, results of P^{32} in two samples and of Be^7 in one sample did not check well. These samples were subjected to additional chemical purification steps, after which the results agreed with each other. Lack of agreement in some of these samples was attributed to contamination by large amounts of fission activities brought into the laboratory by the rain samples. It appears desirable to separate the glassware and the laboratory space used in the preliminary purification steps from those used during final stages.

Experimental results are presented in Table 1. The listed concentrations (atoms/cc) represent the mean of the two extreme values obtained from the three determinations. The errors quoted are based on the observed spread in the determinations. The concentrations of Sr^{90} and Sr^{90} and the ratio of their counting rates at the time of rain are given in Table 2.

The annual fallout rates of the various isotopes, obtained by multiplying the observed average isotope concentrations (atoms/cc) by the average rainfall in Bombay's zonal belt (98 cm), are given in Table 3. In the same table are also given the fallout rates measured at latitudes 10° to $34\frac{1}{2}^\circ$ N for the years 1956 and 1957 [Rama and Zutshi, 1958; Goel and others, 1959; Goel and others, 1957]. The estimated fallout rates [Peters, 1958] for a pure cosmic-ray production of activity produced in the troposphere alone are given in column 5; a value of 30 days was used for the average period of removal of activity from the troposphere.

TABLE 1. Measured Isotope Concentrations in Bombay Rains (1958)

Date	Be ⁷ , atoms/cc	S ³⁵ , atoms/cc	P ³² , atoms/cc	Be ⁷ /P ³²	S ³⁵ /P ³²	S ³⁵ /Be ⁷
June 20	940 ± 300	230 ± 30	11 ± 1	82	20	0.24
June 21	5300 ± 400	1150 ± 100	65 ± 5	82	18	0.22
June 22	4900 ± 400	980 ± 100	67 ± 7	73	14	0.20
June 26	5300 ± 300	540 ± 100	35 ± 5	152	15	0.10
July 1	2400 ± 300	160 ± 20	18 ± 1	133	9	0.07
July 2	4500 ± 400	450 ± 50	25 ± 2	180	18	0.10
July 5	4000 ± 300	440 ± 50	33 ± 3	121	13	0.11
July 7	3650 ± 350	1700 ± 100	62 ± 3	59	26	0.47
July 9	2200 ± 200	1650 ± 50	13 ± 2	169	130	0.75
July 11	1100 ± 200	620 ± 50	20 ± 7	55	31	0.56
July 12	3150 ± 450	1360 ± 50	27 ± 2	115	49	0.43
July 12	4150 ± 450	1650 ± 150	43 ± 3	96	37	0.40
July 20	6400 ± 700	3200 ± 100	50 ± 2	128	64	0.50
July 29	10000 ± 500	3500 ± 150	42 ± 6	238	83	0.35
Aug. 3	8350 ± 250	3450 ± 450	73 ± 5	114	47	0.41
Aug. 5	4670 ± 200	2300 ± 200	46 ± 4	101	50	0.50
Aug. 18	3050 ± 150	1500 ± 200	48 ± 7	63	31	0.49
Aug. 19	5350 ± 300	1850 ± 100	100 ± 20	54	19	0.35
Sept. 11	3150 ± 500	1840 ± 100	55 ± 4	57	34	0.58
average	4350	1500	43.5			

the fallout rates of P³² and Be⁷ are constant with a constant annual deposition each from 1956 to 1958. The agreement is very good for the case of Be⁷ where statistical fluctuations are smaller than for the case of P³². The fallout of S³⁵ is seen to be substantially higher during 1958 and appears to lie outside the range of natural fluctuations.

Interpretation of results. The large-scale fallout of nuclear weapons has resulted in the production of large amounts of various radioactive isotopes in the atmosphere. We are faced with the problem of determining whether the natural concentrations of the isotopes Be⁷, P³², and S³⁵ in rains have been appreciably modified because of artificial production at anytime. The artificial production of P³² and S³⁵ seems to be confined to occur during and following high-yield nuclear detonations on the surface of the sea.² Tests were carried out by the United States during the spring of 1956 and again in the spring of 1958. The fallout data summarized in Table 3 indeed suggest some contribution of artificially produced P³² during 1956 and 1958.

The most likely production reactions are S³²(p, p³²) P³², Cl³⁵(n, p) S³⁵ and S³⁴(n, γ) S³⁵.

However, within the uncertainties in the fallout estimates, its fallout is consistent with its being constant during 1956–1958. The fallout of S³⁵, on the other hand, was appreciably higher during 1956 and 1958 than during 1957. Its concentration in rains compared with Be⁷ during 1958 is so high that it cannot be accounted for only by cosmic ray production. The mean concentrations are 1500 and 4350 atoms/cm³ of S³⁵ and Be⁷, respectively. Since stars produced in argon amount to less than 1 per cent of stars produced in nitrogen and oxygen nuclei, the observed S³⁵ abundance requires that more than 1.7 S³⁵ nuclei be produced in each argon star. (The yield of Be⁷/star in nitrogen and oxygen is taken to be 5 per cent.) Furthermore, the ratio S³⁵/Be⁷ varied by a factor of about 11, whereas, as shown by *Lal and others* [1958], the expected variation due to differences in trajectories of air masses in the troposphere is less than 1.64. These arguments conclusively prove that appreciable amounts of S³⁵ were added to the atmosphere by nuclear weapons tests.

Further proof that artificial production of P³² has not been very significant is provided by a study of the ratio of the concentrations of the

TABLE 2. Sr^{90} and Sr^{89} Concentrations in Bombay Rains (1958)

Sample No.	Sr^{90} , atoms/cc	Sr^{89} , atoms/cc
1	$7.2 \pm 2.0 \times 10^3$	$1.6 \pm 0.5 \times 10^3$
6	$18 \pm 3.0 \times 10^3$	$2.0 \pm 0.3 \times 10^3$
8	$66 \pm 2.6 \times 10^3$	$15.2 \pm 2.3 \times 10^3$
9	$23 \pm 1.7 \times 10^3$	$6.7 \pm 1.6 \times 10^3$
10	$190 \pm 20 \times 10^3$	$8.1 \pm 0.4 \times 10^3$
11	$324 \pm 40 \times 10^3$	$9.2 \pm 1.6 \times 10^3$
12	$35 \pm 1.7 \times 10^3$	$4.6 \pm 0.5 \times 10^3$
13	$460 \pm 20 \times 10^3$	$54 \pm 2.0 \times 10^3$
15	$62 \pm 2.4 \times 10^3$	$9.5 \pm 1.0 \times 10^3$
16	$66 \pm 1.7 \times 10^3$	$10.5 \pm 1.0 \times 10^3$
17	$37 \pm 3.0 \times 10^3$	$6.8 \pm 0.3 \times 10^3$
18	$24 \pm 0.5 \times 10^3$	$14.6 \pm 0.5 \times 10^3$
19	$76 \pm 1.1 \times 10^3$	$13 \pm 2.0 \times 10^3$

$$\text{Mean} \begin{cases} \text{Sr}^{90} = 1.1 \times 10^5 \text{ atoms/cc} \\ \text{Sr}^{89} = 1.4 \times 10^4 \text{ atoms/cc} \end{cases}$$

isotopes Be^7 and P^{32} . The ratio was confined to values lying between ~ 60 and 240 during 1958. Such variations would be expected to arise from the differences in the air-mass trajectories, as discussed by *Lal and others* [1958]. If P^{32} was artificially introduced in substantial amounts at any time we would expect this ratio to vary widely outside this range.

The fallout rate of P^{32} and the minimum ob-

TABLE 3. Annual Deposition Rates of P^{32} , Be^7 , and S^{35} at Tropical Latitudes (1956-1958)

Radio-isotope	Observed Annual Deposition Rates			Estimated Annual Deposition Rates, atoms/cm ² year
	1956	1957	1958	
P^{32}	3.1×10^3 (10)*	2.7×10^3 (30)	4.3×10^3 (19)	1.6×10^3
Be^7	4.8×10^5 (76)	4.0×10^5 (75)	4.3×10^5 (19)	4.7×10^5
S^{35}	1.3×10^5 (9)	3.1×10^4 (10)	1.5×10^5 (19)	6.2×10^3

* Numbers in parentheses below each figure represent the number of samples on which the figure is based.

served ratio of $\text{Be}^7/\text{P}^{32}$ require a higher production rate of P^{32} compared with that calculated by Lal and others by a factor of about

Average renewal period of activity from troposphere. We shall confine our study to the Be^7 and P^{32} data, since only their activity is seen to be chiefly of cosmic-ray origin.

It was pointed out by *Goel and others* [1958] that the absolute concentrations of the radioisotope Be^7 showed very large variations from one rain to another (by factors of up to 4) while the ratios of $\text{Be}^7/\text{P}^{32}$ remained confined to values differing by a factor of ~ 4 . In the rain samples on which we report in this paper we found Be^7 concentrations (atoms/cc) between ~ 1000 and $10,000$; the $\text{Be}^7/\text{P}^{32}$ ratios range from ~ 60 to 240 (Table 1). Such behavior was expected, since the ratios of isotope concentrations do not depend on the variations in moisture condensed per unit mass of air. *Goel and others* [1958] have calculated the expected ratios of isotopic concentrations under certain simplifying assumptions regarding the circulation of air in the troposphere. It was assumed that air is instantaneously cleaned of its radioactivity after a condensation leading to precipitation. Furthermore, convection in the troposphere was assumed to be sufficiently strong so that the production rates in the troposphere, which vary exponentially with a mean free path of $\sim 150 \text{ g cm}^{-2}$, can be approximated by an average value. The ratio of the concentrations C_1 , C_2 of two isotopes of mean life τ_1 and τ_2 at any time t after a cleansing by condensation is then given by

$$\frac{C_1}{C_2} = \frac{P_1 \tau_1}{P_2 \tau_2} \frac{1 - e^{-t/\tau_1}}{1 - e^{-t/\tau_2}}$$

the two isotopes per gram of air in the troposphere.

P_1 and P_2 are the average production rates in the troposphere. For small values of irradiation time t the ratio C_1/C_2 becomes equal to P_1/P_2 . This is for air circulating entirely within the troposphere, the ratio $\text{Be}^7/\text{P}^{32}$ is expected to range from $P_{\text{Be}^7}/P_{\text{P}^{32}}$ for short irradiation to $(P_{\text{Be}^7}/P_{\text{P}^{32}})e^{t(\tau_1 - \tau_2)}$ for air which is irradiated for periods that are long compared with the half-lives of these isotopes. Stratospheric air will have the maximum ratio, namely $3.7(P_{\text{Be}^7}/P_{\text{P}^{32}})$, before its descent into the troposphere. After such an air mass has descended into the troposphere

ratio will increase steeply with time for periods up to ~ 3 months due to the preferential decay of the short-lived isotope P^{33} . Only then can the stratospheric activity have been reduced to a level where the much lower isotope-production rates in the troposphere become important.

The most probable lowest value of the ratio $\text{Be}^7/\text{P}^{33}$ during the 1958 monsoon is seen to be ± 5 (Table 1). The highest value during the same period is 238 ± 35 , corresponding to a range of 3.9 ± 0.7 . During 1957⁸ the highest value was smaller than 220 in each case except for one sample where the value was 290 ± 45 . Our results therefore indicate that we are dealing with activity derived from air irradiated in the troposphere.

We may then calculate the period t in each sample, using equation (1) and the lowest observed ratio, 60 for P_1/P_2 . The calculated periods t , which represent the time of irradiation in a troposphere which is assumed to be well mixed, are seen to follow an exponential distribution with an average period of 45 days (Fig. 1). This value is somewhat higher than, but consistent with the value of 35 days derived last year [Goel and others, 1959]. The two measurements are consistent with a value of 40 ± 5 days. Our value does not disagree with the value of 1 month observed by Stewart and others [1957] for fission products at middle latitudes. This value is actually a lower limit, since lateral zonal spread of the fission products has not been considered. It may be of some interest to point out that the residence period of water in the troposphere, the removal of which occurs chiefly through precipitation, is found to be 40 days [Junge and Werby, 1958]. The residence period of bulk water in the troposphere is an order of magnitude smaller than the removal period, as can be seen from the total annual precipitation and the average water content of the atmosphere. This seems to imply that the period of 40 days represents the

The minimum value of the ratio $\text{Be}^7/\text{P}^{33}$, based on 11 samples in 1957, was 100 ± 10 . It is somewhat higher than the value 60 ± 5 among 19 samples in 1958, but may not be inconsistent in view of the small number of samples analyzed. If the difference is real, it indicates some bomb-produced P^{33} in rains in 1958.

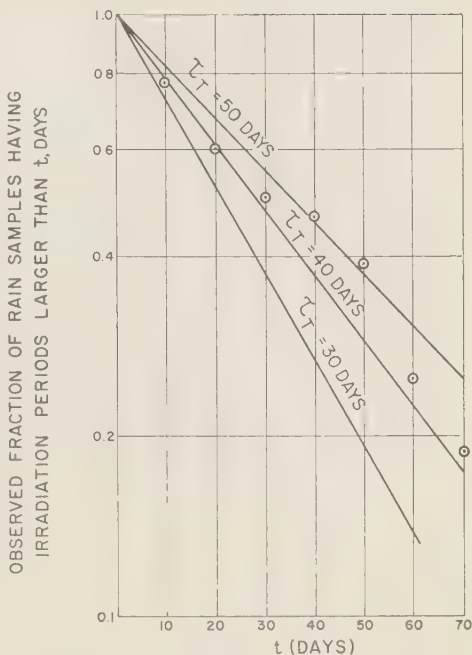


Fig. 1. The integral frequency distribution of irradiation periods, based on the observed ratio $\text{Be}^7/\text{P}^{33}$, in 19 rain samples. The solid lines show the expected behavior for values of 30, 40, and 50 days for the average removal period of activity from the troposphere.

time which air spends in different regions of the troposphere before picking up moisture at lower altitudes in the atmosphere.

Acknowledgments. We are grateful to B. Peters and M. G. K. Menon for encouragement and discussions, and to E. Noronha for skilled assistance in the analysis of samples.

Our thanks are due Shankar Das and S. B. Hingorani of the Atomic Energy Establishment, Bombay, for advice in connection with the extraction of strontium.

We are indebted to E. A. Martell for his valuable comments on artificial production of P^{33} and S^{35} .

REFERENCES

- Goel, P. S., N. Narsappaya, C. Prabhakara, Rama, and P. K. Zutshi, Study of cosmic ray produced short-lived isotopes P^{33} , P^{32} , Be^7 , and S^{35} in tropical latitudes, *Tellus*, 11, 91-100, 1959.

- Goel, P. S., Rama, and P. K. Zutshi, Symposium on cosmic rays, *Indian J. Meteorol. Geophys.*, **8**, 329-330, 1957.
- Junge, C. E., and R. T. Werby, The concentration of chloride, sodium, potassium, calcium and sulfate in rain water over the United States, *J. Meteorol.*, **15**, 417-425, 1958.
- Lal, D., P. K. Malhotra, and B. Peters, On the production of radioisotopes in the atmosphere by cosmic radiation and their application to meteorology, *J. Atmospheric and Terrest. Phys.*, **12**, 306-328, 1958.
- Osmond, R. G., A. G. Pratchett, and J. B. War-ricker, The determination of long lived fall-out in rain water, *Atomic Energy Research Establ., Gt. Brit., Publ. no. A.E.R.E. C/R 2165*, 1959.
- Peters, B., Cosmic ray produced radioactive isotopes as tracers for studying large-scale atmospheric circulation, *J. Atmospheric and Terrest. Phys.*, **13**, 351-370, 1959.
- Rama and P. K. Zutshi, Annual deposition of cosmic ray produced Be^7 at equatorial latitude, *Tellus*, **10**, 99-103, 1958.
- Stewart, N. G., R. G. D. Osmond, R. N. Crook, and E. M. Fisher, The world-wide deposition of long-lived fission products from nuclear test explosions, *Atomic Energy Research Establ., Gt. Brit., Publ. no. A.E.R.E. HP/R 2354*, 1958.

(Manuscript received October 16, 1959.)

Free Oscillations of the Earth¹

I. Toroidal Oscillations

FREEMAN GILBERT AND GORDON J. F. MACDONALD

*Institute of Geophysics, University of California
Los Angeles 24, California*

Abstract. The free periods of toroidal oscillations of the earth have been computed for two earth models. The lowest period for the Gutenberg model earth is 2651 sec and for the Jeffreys-Bullen model 2732 sec. The surface amplitudes of the oscillations have been computed for three kinds of delta function stress sources—a unit force, a unit couple, and a unit torque—at depths of 600, 250, 100, and 30 km. The amplitudes decrease with increasing depth of the source. For a unit couple at 600 km the maximum amplitude of the lowest period for the Gutenberg model is 1.59×10^{-26} cm, and for the Jeffreys-Bullen model it is 0.70×10^{-26} cm. By using the free periods of oscillation we have extended Love wave phase velocity and group velocity dispersion curves to include long-period Love waves.

The method used to compute the periods and amplitudes of the free oscillations is an extension of the Thomson-Haskell matrix method used in plane layered media.

An example is presented to show the correspondence between the free oscillations and ray theory.

Introduction. The fundamental problem in earthquake seismology is to proceed from scattered observations of earthquakes to a detailed specification of the distribution of velocity within the earth and to descriptions of the disturbances at the source of the earthquake. The customary interpretation is based on ray theory. Ray theory has been remarkably successful in determining the structure of the earth. An alternative approach is based on normal mode theory. In the mathematics of normal mode theory an earthquake is described as a disturbance excited by some initial distortion. The solution represents the propagation of the initial distortion and is written in terms of the normal modes of the earth where the normal modes are functions of space and time. The excitation of a single normal mode would correspond to a motion in which each particle within the earth was either at rest or executing a simple harmonic motion. The particles at rest lie on nodal surfaces. The geometric distribution of the nodal surfaces is fixed by the order of the normal mode. The frequencies of the normal modes are the resonant frequencies of the earth. At these resonant frequencies there is an equal partition of kinetic and potential energy over one or more cycles of oscillation. Because the earth is finite in

size the frequencies of free oscillations are discrete, although they are infinite in number. Of basic importance is the fact that any vibrations excited by initial stresses can be expressed as the sum of natural resonant modes.

The ray-wave theory and the normal mode theory are complementary, and where one is cumbersome the other may lead to useful results. Whereas the techniques of ray interpretation have been highly developed, the normal mode analysis has only recently been applied to seismological problems. Observational studies have emphasized the identification of a particular mode of oscillation [Benioff, 1954; Benioff and others, 1959]. It is our purpose to explore the usefulness of the normal mode approach to seismological problems. In particular, we hope to determine whether the normal mode approach can be used to reduce the number of hypothesized earth models derived from seismic observations as interpreted by ray theory. The distribution of elastic properties and density within the earth determines the resonant frequencies. A comparison of the resonant frequencies of models with those of the earth may provide further restrictions on the internal constitution of the earth. The amplitudes of the various modes of oscillation are fixed by the source characteristics, geometry, intensity, and time history. It thus may be possible to describe more accurately

¹ Institute of Geophysics Publication No. 160.

the source phenomena of earthquakes if the observations are interpreted in terms of the normal mode theory.

As will be shown, the applicability of the normal mode approach depends to a large extent on the development of seismic recording instruments of high resolution at frequencies on the order of 1 to 40 cycles/hour. It is in this range of frequencies that the lines making up the spectrum are sufficiently separated so that a detailed analysis is possible.

The normal mode approach began with the work of *Jaerisch* [1880] who carried out an early investigation of the free oscillations of a homogeneous sphere. He considered radial vibrations (pure compression) and toroidal vibrations (no radial motion), the latter corresponding to SH waves. The first comprehensive treatment was that of *Lamb* [1882] who considered a non-gravitating compressible sphere and discussed both toroidal vibrations (vibrations of the first class) and vibrations involving coupling between tangential and radial displacements (vibrations of the second class). *Jeans* [1903] pointed out that strains due to gravitational forces would be of the same order of magnitude as those produced by elastic restoring forces, but he was unable to take account of gravitation satisfactorily. A method of treating gravitational effects was suggested by *Rayleigh* [1906] and later used by *Love* [1911] and *Jeans* [1923]. *Jeans* showed that the system of free oscillations excited by an earthquake source could be regarded as a system of dispersive surface waves, and he deduced the relation between phase velocity, frequency, and order of spherical harmonic representing a particular mode of oscillation. He was able to show that there were, in addition to the surface waves, other waves diffusing into the earth's interior. In effect *Jeans* showed the correspondence between normal mode theory and ray-wave theory in an elegant way. His paper represents the most significant contribution to the problem since *Lamb's* [1882] study.

The extension of the results for a homogeneous sphere to an inhomogeneous earth model involves formidable computational problems. *Stoney* [1926] suggested a variational method, and most of the work since 1926 has depended on a variational approach. The most thorough treatments

using that approach are those of *Pekeris and Jarosch* [1958] and *Takeuchi* [1959]. In the former only the spheroidal modes were considered and in the latter the toroidal mode. Other applications of the variational method have been made by *Takeuchi* [1950], *Molodenskiy* [1953], *Matumoto and Sato* [1954], *Jobert* [1957], and *Gilbert and MacDonald* [1959]. There are serious difficulties with the variational method: it yields poor estimates of the high frequency free oscillations, and, as a result, we have discarded it in favor of a more direct method.

Alterman, Jarosch and Pekeris [1959] determined the toroidal and spheroidal free oscillations by numerically integrating on an electronic computer the differential equations of motion for an inhomogeneous sphere. They not only computed the free modes for several earth models, but also computed the displacement field due to a compressional point source and represented it as the superposition of free oscillations. *Ben-Menahem* [1959] determined some of the toroidal oscillations by expressing the rigidity and shear velocity in analytic form and solving the equations of motion directly.

In this paper we use a third approach based on methods of matrix manipulation. The technique was formulated by *Thomson* [1950] and extended by *Haskell* [1953] to compute dispersion curves for a plane stratified medium. *Haskell's* solution has been programmed for an IBM 650 computer by *Dorman* [1959], and he has obtained dispersion curves for a variety of velocity and density profiles. In the present study we have extended the Thomson-Haskell formulation to a layered sphere and carried out the numerical computations on two models of the earth.

In the Thomson-Haskell method the known solution for an elastic spherical shell is used. The earth is assumed to be made up of a number of concentric, spherical, elastic shells. The solutions for each shell are continued by use of the appropriate boundary conditions. The utilization of the exact solution for a given shell does not involve many of the error problems found in the direct numerical integration of a second order differential equation.

Formulation of the problem. Consider the earth to be a layered sphere, each layer being homogeneous, isotropic, and perfectly elastic. In the

present paper we consider only motions parallel to the spherical surfaces and require that the vector potential of the motion be radial. Such motions are known as toroidal motions and correspond to SH waves. If $\mathbf{U}(\mathbf{r})$ is the Fourier transform with respect to time of the displacement vector $\mathbf{u}(\mathbf{r}, t)$

$$\mathbf{U}(\mathbf{r}) = \int_{-\infty}^{\infty} \mathbf{u}(\mathbf{r}, t) e^{i\omega t} dt \quad (1)$$

then

$$\begin{aligned} \mathbf{U}(\mathbf{r}) = \sum_{l=1}^{\infty} \sum_{m=0}^l \sqrt{l(l+1)} \mathbf{C}_{ml}^{\sigma}(\theta, \varphi) \\ \cdot [A_{ml}^{\sigma} j_l(kr) + B_{ml}^{\sigma} y_l(kr)] \end{aligned} \quad (2)$$

[Morse and Feshbach, 1953, chap. 13]. The subscripts m and l refer to the degree and order of the spherical harmonic. The superscript σ is either e or o , depending on whether we use $\cos(m\varphi)$ or $\sin(m\varphi)$ in the spherical harmonic. μ is rigidity, and $k = \omega/c$, where c is the shear wave velocity. The radial functions, j_l and y_l are spherical Bessel functions of the first and second kind. The coefficients A , B are determined from the boundary conditions and \mathbf{C} is a vector surface harmonic. The radial traction is given by

$$\begin{aligned} T_r = \sum_{l=1}^{\infty} \sum_{m=0}^l \sqrt{l(l+1)} \mathbf{C}_{ml}^{\sigma}(\theta, \varphi) \\ \cdot \{ \mu k A_{ml}^{\sigma} [j_l'(kr) - (kr)^{-1} j_l(kr)] \\ + \mu k B_{ml}^{\sigma} [y_l'(kr) - (kr)^{-1} y_l(kr)] \}. \end{aligned} \quad (3)$$

For $l = 1$, equation (2) gives the simplest angular dependence for the vibration of a sphere. In this case the displacement is such that the interior of the sphere rotates about a given axis in one direction, while the outer part of the sphere rotates in the opposite direction about the same axis. The modes for $l = 2$ are such that the Northern Hemisphere rotates in one direction about the preferred axis and the Southern Hemisphere rotates in the other direction, with the equatorial plane being a nodal surface.

If we denote the bracketed terms in (2) and (3) by U and T , respectively, then

$$\begin{bmatrix} U \\ T \end{bmatrix} = \begin{bmatrix} j_l(kr) \\ \mu k [j_l'(kr) - (kr)^{-1} j_l(kr)] \end{bmatrix}$$

where the superscripts and subscripts have been suppressed on A and B . Let n be the subscript used to denote the n th spherical shell, let r_n be the radius of the bottom of the n th shell, and let r_{n+1} be the radius of the top of the shell. The radius of the outer surface is $r = r_{N+1} = a$, and the radius of the core-mantle boundary is $r = r_1 = b$. In considering the toroidal oscillations we can neglect the fluid core and obtain the oscillations of the mantle alone.

Equation (4) may be rewritten with the aid of the subscript n as

$$\begin{bmatrix} U_n(r) \\ T_n(r) \end{bmatrix} = \mathcal{G}_n(r) \cdot \begin{bmatrix} A_n \\ B_n \end{bmatrix} \quad (5)$$

where \mathcal{G}_n is the 2×2 matrix in (4). \mathcal{G}_n is the function of the density and rigidity of the layer whose inner radius is r_n . The inverse of (5) is

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \mathcal{G}_n^{-1}(r) \cdot \begin{bmatrix} U_n(r) \\ T_n(r) \end{bmatrix} \quad (6)$$

provided \mathcal{G}_n is not singular. Combining (5) and (6), we relate the stress and displacement at the level r_n to the same quantities at r_{n+1} ;

$$\begin{bmatrix} U_n(r_n) \\ T_n(r_n) \end{bmatrix} = \mathcal{G}_n(r_n) \mathcal{G}_n^{-1}(r_{n+1}) \begin{bmatrix} U_n(r_{n+1}) \\ T_n(r_{n+1}) \end{bmatrix} \quad (7)$$

The notation

$$\mathcal{G}_n(r_n) \mathcal{G}_n^{-1}(r_{n+1}) = \mathcal{B}_n \quad (8)$$

$$\mathcal{B}_1 \mathcal{B}_2 \cdots \mathcal{B}_n = \mathcal{C}^n \quad (9)$$

is convenient for expressing the boundary conditions. Boundary conditions require continuity of U and T so that the multilayer relationship

$$\begin{bmatrix} U_1(r_1) \\ T_1(r_1) \end{bmatrix} = \mathcal{C}^N \begin{bmatrix} U_N(r_{N+1}) \\ T_N(r_{N+1}) \end{bmatrix} \quad (10)$$

can easily be obtained. In terms of the inner and outer radius of the mantle we have

$$\begin{bmatrix} U_1(b) \\ T_1(b) \end{bmatrix} = \mathcal{C}^N \begin{bmatrix} U_N(a) \\ T_N(a) \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} y_l(kr) \\ \mu k [y_l'(kr) - (kr)^{-1} y_l(kr)] \end{bmatrix} \cdot \begin{bmatrix} A \\ B \end{bmatrix} \quad (4)$$

The stress must vanish at the outer surface $r = a$. The stress is taken to vanish at the core-mantle boundary. We thus assume that the stresses associated with the viscosity of the core are small compared with the rigidity of the mantle. Equation (11) becomes

$$\begin{bmatrix} U_1(b) \\ 0 \end{bmatrix} = \mathcal{C}^N \begin{bmatrix} U_N(a) \\ 0 \end{bmatrix} \quad (12)$$

or

$$\begin{aligned} U_1(b) &= \mathcal{C}_{11}^N U_N(a) \\ 0 &= \mathcal{C}_{21}^N U_N(a) \end{aligned} \quad (13)$$

$$\therefore \mathcal{C}_{21}^N = 0$$

where the subscripts on \mathcal{C} denote the element of the matrix. The roots of

$$\mathcal{C}_{21}^N = 0 \quad (14)$$

are the resonant frequencies for toroidal oscillations. We denote the n th root for the l th harmonic of the toroidal oscillations by ${}_n T_l$. The determination of the frequencies of the normal modes reduces to that of evaluating the matrix given in (4) for each layer, carrying out the matrix multiplications given by (8) and (9) and finding the zeros of the function \mathcal{C}_{21}^N .

We consider a stress source at a radius $r = r_{p+1}$ of

$$T_p(r_{p+1}) - T_{p+1}(r_{p+1}) = \mathfrak{J} \quad (15)$$

We take the source as a delta function at $\theta = \theta_0$, $\varphi = 0$ pointed in the θ direction

$$\sqrt{l(l+1)} \mathbf{C}_{m1}^\sigma(\theta, \varphi) \mathfrak{J} = \frac{\hat{\theta} \delta(\theta - \theta_0) \delta(\varphi)}{r_{p+1}^2 \sin \theta_0} \quad (16)$$

for the harmonic of order l and degree m . We take the product of (16) with \mathbf{C}_{m1}^σ and integrate over the spherical surface. Using the orthogonality relation for \mathbf{C}_{m1}^σ [Morse and Feshbach, 1953, pp. 1899-1900] we find $\sigma = o$ (odd) and

$$\mathfrak{J} = \frac{\epsilon_m(l-m)! (2l+1) m P_l^m(\cos \theta_0)}{4\pi(l+m)! r_{p+1}^2 \sin \theta_0 l(l+1)} \quad (17)$$

where ϵ_m is 1 for $m = 0$ and 2 for $m > 0$. We now let θ_0 approach 0. The source is then at the pole of the coordinate system. The limit

of (17) is 0 unless $m = 1$. For $m = 1$ we have

$$\lim_{\theta_0 \rightarrow 0} \frac{P_l^1(\cos \theta_0)}{\sin \theta_0} = \frac{l(l+1)}{2} \quad (18)$$

Equation (17) then becomes

$$\mathfrak{J} = \frac{(2l+1)}{4\pi r_{p+1}^2 l(l+1)} \quad (19)$$

The boundary conditions in the matrix formulation become

$$\begin{aligned} \begin{bmatrix} U_1 \\ 0 \end{bmatrix} &= \mathcal{C}^p \begin{bmatrix} U_p(r_{p+1}) \\ T_p(r_{p+1}) \end{bmatrix} \\ &= \mathcal{C}^p \begin{bmatrix} U_{p+1}(r_{p+1}) \\ T_{p+1}(r_{p+1}) + \mathfrak{J} \end{bmatrix} \\ &= \mathcal{C}^p \begin{bmatrix} U_{p+1}(r_{p+1}) \\ T_{p+1}(r_{p+1}) \end{bmatrix} + \mathcal{C}^p \begin{bmatrix} 0 \\ \mathfrak{J} \end{bmatrix} \\ &= \mathcal{C}^N \begin{bmatrix} U_N(a) \\ 0 \end{bmatrix} + \mathcal{C}^p \begin{bmatrix} 0 \\ \mathfrak{J} \end{bmatrix} \end{aligned} \quad (20)$$

The displacement at the surface, $r = a$, is given by

$$U_N(a) = -\mathcal{C}_{22}^p \mathfrak{J} / \mathcal{C}_{21}^N \quad (21)$$

In (21) the displacement is a function of frequency. To obtain the time response for a delta function source we use the Fourier inversion of (21),

$$u_{N1}(a, t) = -2 \sum_{n=1}^{\infty} \mathcal{C}_{22}^p({}_n \omega_l) \mathfrak{J} \cdot \sin({}_n \omega_l t) / \mathcal{C}_{21}^{N'}({}_n \omega_l) \quad (22)$$

where the sum in (22) is the time response for each l in the spherical harmonic series. For any time behavior other than a delta function (22) can be convolved with the specified time behavior of the source to obtain the time behavior of the response. Thus the variation of displacement with time and position on the surface can be calculated, given the time history of the source and the spatial distribution of the source. Alternatively, the frequency response at a given station can be calculated by multiplying the input spectrum at the source with the transfer function for the earth, where the transfer function is defined by the ratio

$$-C_{22}^P(\omega)/C_{21}^N(\omega) \quad (23)$$

The factor \mathfrak{J} for a couple is the same as that given in (19) for a unit force, but the displacement distribution for the couple is the derivative of the displacement for the force taken in the spherical surface normal to the force direction. The couple is obtained by applying two forces of equal strengths pointing in opposite directions. The strength of the couple is the product of the force strength and the distance between them. As this distance becomes small the limit approached is the derivative of the distribution for a single force.

The unit force and the couple excite both toroidal oscillations and spheroidal oscillations, but in this paper we consider only the former. A source that excites only toroidal oscillations is a 'torque' source, a circulation-type source with motion only in the spherical surfaces of the earth. We consider a torque source located at the pole of the coordinate system so that the motion is in the φ direction only and is a function of r and θ . In this case we replace the vector $\sqrt{l(l+1)} C_{mi}^*(\theta, \varphi)$ by the vector $P_l(\cos \theta)$ in (2) and (3). Then in place of (16) we have

$$(\cos \theta)\mathfrak{J} = \delta(\theta - \theta_0)/r_{p+1}^2 \sin \theta_0, \quad \theta_0 \rightarrow 0 \quad (24)$$

Using the orthogonality properties of the P_l , we find that for a torque source

$$\mathfrak{J} = (2l+1)P_l(\cos \theta_0)/4\pi r_{p+1}^2 \quad (25)$$

we let $\theta_0 \rightarrow 0$ in order to place the source at the pole so that

$$\mathfrak{J} = (2l+1)/4\pi r_{p+1}^2 \quad (26)$$

for a torque source at the pole.

The displacement patterns on the surface of the earth are shown in Figures 1 to 4 for the first four harmonics for a couple source at the pole.

Model earth. In order to calculate the resonant frequencies of the earth, we require the distribution of the rigidity in the mantle of the earth. The ratio of the rigidity to the density can be obtained directly from curves of S-wave velocity versus depth. The curves of S-wave velocity are derived from seismic observation

by use of ray theory with the direct integration of the travel-time curves. At present two detailed interpretations have been carried out [Jeffreys, 1959; Gutenberg, 1953]. The major features of both distributions are similar, but Gutenberg's solution differs in a significant way for the upper part of the mantle. Gutenberg's proposal is that at a depth of about 100 km the

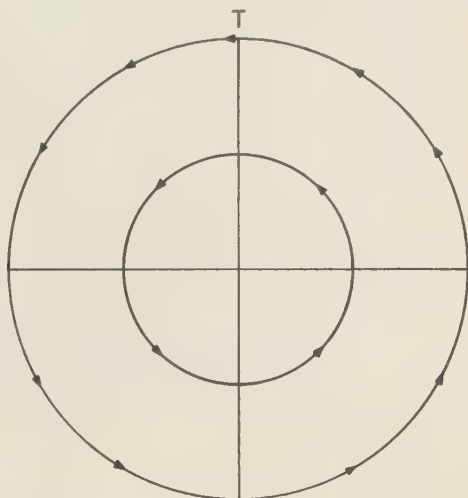


Fig. 1. Displacement pattern on the surface of the earth for a couple source at the pole; $l = 1$.

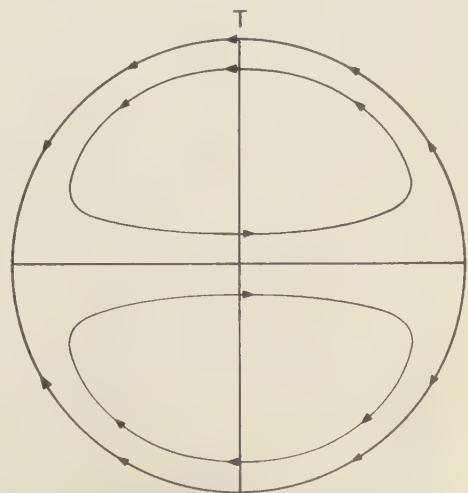


Fig. 2. Displacement pattern on the surface of the earth for a couple source at the pole; $l = 2$.

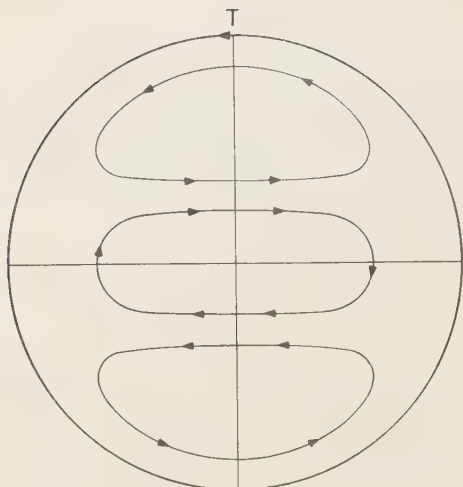


Fig. 3. Displacement pattern on the surface of the earth for a couple source at the pole; $l = 3$.

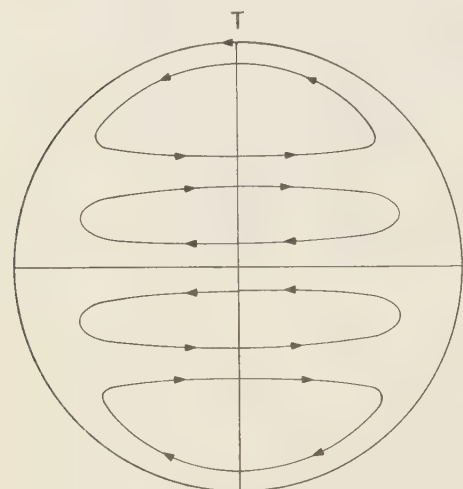


Fig. 4. Displacement pattern on the surface of the earth for a couple source at the pole; $l = 4$.

S-wave velocity decreases; thereafter it increases again and nearly coincides with Jeffreys' values at depths greater than about 600 km. The difference in interpretation by Jeffreys and Gutenberg of rather similar data is connected with the identification of various phases of a seismogram. Recent studies of surface waves [Dorman, 1959] appear to substantiate Gutenberg's hypothesis of a low-velocity layer,

Since the toroidal oscillations involve both the rigidity and the S-wave velocity, it is necessary to assume a distribution of density within the mantle in order to obtain the rigidity from the distribution of S-wave velocity. Bullen [1953] obtained the distribution of density by using Jeffreys' solution for the velocity-depth relation coupled with the mean density of the earth and the moment of inertia. Bullen has given a number of alternative distributions of densities depending on different assumptions concerning the discontinuities at the boundary of the core mantle and at the inner core. Bullard [1957] carried out a study of allowable density distributions and concluded that Bullen's solution having an inhomogeneous layer between the depths of 413 and 984 km, should not be in error by more than 0.3 g/cm³ in the mantle provided the temperature distribution is adiabatic. In the present work we shall adopt Bullen's density distribution within the mantle and use both Gutenberg's and Jeffreys' distribution of velocities. The rigidity within the mantle is then obtained by combining the distribution of density and S-wave velocity.

In both the Jeffreys and Gutenberg models of the earths the mantle is approximated by 30 concentric spherical shells. In both models a crust of 30 km having a low velocity is assumed. The constants used as input data for the calculations are listed in Table 1.

The mass average properties of the Gutenberg model earth and of the Jeffreys-Bullen model earth are given in Table 2. It is seen that the mass average properties of the two models are almost identical. Any differences in the resonant frequencies of the two models result solely from the distribution of rigidity and density within the models and not from their average properties. In Table 2 the mass average properties of model α and model β treated by Alterman, and others [1959] are also listed. Model α consists of a homogeneous, solid mantle enclosing a homogeneous, liquid core with the properties of the mantle equal to the average properties of the mantle of Bullen's model B. Model β is a homogeneous solid with properties equal to the average over the whole globe of Bullen's model B.

Numerical computations. The computations of the resonant frequencies and of the amplitudes of the free oscillations were carried out on the

TABLE 1. Description of Models

Radii Bounding Layers, km $\times 10^3$	Gutenberg		Jeffreys-Bullen		Radii Bounding Layers, km $\times 10^3$	Gutenberg		Jeffreys-Bullen	
	V_s , km/sec	μ , dynes/cm ² $\times 10^{12}$	V_s , km/sec	μ , dynes/cm ² $\times 10^{12}$		V_s , km/sec	μ , dynes/cm ² $\times 10^{12}$	V_s , km/sec	μ , dynes/cm ² $\times 10^{12}$
6.37	3.60	0.369	3.60	0.369	5.32	6.44	1.95	6.44	1.95
6.34	4.43	0.656	4.40	0.646	5.20	6.52	2.04	6.52	2.04
6.27	4.38	0.651	4.49	0.684	5.07	6.60	2.12	6.60	2.12
6.21	4.40	0.669	4.59	0.729	4.94	6.68	2.20	6.68	2.20
6.15	4.51	0.714	4.69	0.773	4.82	6.76	2.28	6.74	2.28
6.08	4.66	0.772	4.80	0.819	4.69	6.82	2.36	6.80	2.35
6.02	4.87	0.855	4.91	0.869	4.56	6.89	2.44	6.86	2.42
5.96	5.12	0.968	5.10	0.963	4.44	6.95	2.52	6.92	2.50
5.89	5.34	1.10	5.35	1.11	4.31	6.99	2.58	6.98	2.57
5.83	5.52	1.23	5.57	1.25	4.18	7.02	2.63	7.04	2.65
5.77	5.71	1.37	5.76	1.39	4.06	7.09	2.71	7.11	2.75
5.70	5.91	1.50	5.93	1.51	3.93	7.16	2.80	7.17	2.81
5.64	6.06	1.62	6.07	1.62	3.80	7.21	2.87	7.23	2.89
5.58	6.19	1.72	6.18	1.72	3.68	7.27	2.93	7.29	2.97
5.51	6.28	1.80	6.26	1.79	3.55	7.24	2.95	7.31	3.01
5.45	6.35	1.87	6.35	1.87	3.49				

M 709 computer of the Western Data Processing Center, University of California at Los Angeles. The calculations were designed to evaluate the matrix elements given in equation (9) for each layer and to carry out the matrix manipulations indicated in equations (8) to (11). Subroutines were written to evaluate the spherical Bessel functions of the first and second kind. An analysis of the error involved in the calculations shows that at least four significant figures are carried in the calculation of the resonant frequencies for oscillations of order 15 and less. In the higher-order oscillations the rapid increase in magnitude of the Bessel functions for arguments of small order increases the error in the calculation. Only two significant

figures are carried in the calculation of the resonant frequency for the spherical oscillation of order 18.

It would be desirable to extend the calculation

TABLE 2. Mass Average Properties of Models

Model	μ , dynes/cm ² $\times 10^{12}$	V_s , km/sec
Gutenberg	1.75	6.06
Jeffreys-Bullen	1.75	6.03
Model α^*	1.74	6.24
Model β^*	1.46	5.14

* Alterman, Jarosch, and Pekeris [1959].

TABLE 3. Resonant Frequencies of the Earth for Toroidal Oscillations

Oscillation	Gutenberg		Jeffreys-Bullen		Oscillation	Gutenberg		Jeffreys-Bullen	
	Frequency, cycles/hour	Period, sec	Frequency, cycles/hour	Period, sec		Frequency, cycles/hour	Period, sec	Frequency, cycles/hour	Period, sec
${}_2T_1$	4.49	801	4.47	806	${}_1T_{10}$	5.83	617	5.84	608
${}_3T_1$	7.96	452	8.13	443	${}_2T_{10}$	9.47	380	9.42	381
${}_4T_1$	11.65	309	11.73	307	${}_3T_{10}$	11.88	303	11.88	303
					${}_4T_{10}$	14.46	249	14.52	249
${}_1T_2$	1.36	2651	1.32	2732	${}_1T_{11}$	6.29	572	6.30	572
${}_2T_2$	4.79	751	4.81	748	${}_2T_{11}$	10.08	357	10.03	357
${}_3T_2$	8.16	441	8.16	441	${}_3T_{11}$	12.54	287	12.50	287
${}_4T_2$	11.76	306	11.84	304	${}_4T_{11}$	15.00	240	15.06	239
${}_1T_3$	2.12	1694	2.09	1721	${}_1T_{12}$	6.72	536	6.74	536
${}_2T_3$	5.24	687	5.25	686	${}_2T_{12}$	10.65	338	10.65	338
${}_3T_3$	8.37	430	8.41	428	${}_3T_{12}$	13.19	273	13.19	273
${}_4T_3$	11.92	302	11.96	301	${}_4T_{12}$	15.58	231	15.65	231
${}_1T_4$	2.78	1295	2.84	1269	${}_1T_{13}$	7.16	503	7.20	503
${}_2T_4$	5.76	625	5.79	622	${}_2T_{13}$	11.21	321	11.21	321
${}_3T_4$	8.65	416	8.65	416	${}_3T_{13}$	13.90	259	13.90	259
${}_4T_4$	12.16	296	12.20	295	${}_4T_{13}$	16.22	222	16.29	222
${}_1T_5$	3.36	1072	3.35	1074	${}_1T_{14}$	7.56	476	7.66	476
${}_2T_5$	6.33	569	6.24	577	${}_2T_{14}$	11.76	306	11.84	306
${}_3T_5$	9.04	398	9.02	399	${}_3T_{14}$	14.52	248	14.57	248
${}_4T_5$	12.41	290	12.41	290	${}_4T_{14}$	16.90	213	16.90	213
${}_1T_6$	3.92	918	3.94	913	${}_1T_{15}$	7.98	451	8.08	451
${}_2T_6$	7.02	513	6.96	517	${}_2T_{15}$	12.29	293	12.33	293
${}_3T_6$	9.47	380	9.55	377	${}_3T_{15}$	15.19	237	15.25	237
${}_4T_6$	12.77	282	12.81	281	${}_4T_{15}$	17.56	205	17.65	205
${}_1T_7$	4.43	812	4.47	806	${}_1T_{16}$	8.43	427	8.43	427
${}_2T_7$	7.63	472	7.56	476	${}_2T_{16}$	12.90	279	13.09	279
${}_3T_7$	10.03	359	10.03	359	${}_3T_{16}$	15.79	228	15.93	228
${}_4T_7$	13.09	275	13.14	274	${}_4T_{16}$	18.27	197	18.37	197
${}_1T_8$	4.92	732	4.97	725	${}_1T_{17}$	8.80	409	8.74	409
${}_2T_8$	8.26	436	8.26	436	${}_2T_{17}$	13.38	269	13.48	269
${}_3T_8$	10.40	340	10.62	339	${}_3T_{17}$	16.44	219	16.51	219
${}_4T_8$	13.53	266	13.53	266	${}_4T_{17}$	18.95	190	18.95	190
${}_1T_9$	5.39	668	5.45	661	${}_1T_{18}$	9.18	392	8.98	400
${}_2T_9$	8.84	407	8.89	405	${}_2T_{18}$	13.95	258	14.06	258
${}_3T_9$	11.21	321	11.21	321	${}_3T_{18}$	17.00	211	17.14	211
${}_4T_9$	13.95	258	14.01	257	${}_4T_{18}$	19.67	183	19.56	183

tions to oscillations of higher order for a direct comparison with calculations of the modes of a flat, layered earth. However, this would require reprogramming the calculations in double precision arithmetic.

Resonant frequencies of the earth for toroidal oscillations. The resonant frequencies for toroi-

dal oscillations of the Gutenberg and Jeffreys-Bullen model earths are listed in Table 3. In this table the notation ${}_nT_l$ is used to identify the various oscillations. The l refers to the order of the oscillation; the n indicates whether the oscillation is a fundamental or an overtone. The oscillation of lowest frequency is ${}_1T_2$ with

TABLE 4. Comparison of Periods of Toroidal Oscillations

Oscillation	Gutenberg sec	Jeffreys-Bullen sec	Bullen model B after <i>Alterman and others</i> [1959] sec	Bullen, after <i>Takeuchi</i> [1959] sec
T_2	2651	2732	2646	2606
T_1	751	748	762	...
T_2	441	441	438	...
T_3	1694	1721	1716	...
T_4	687	686	696	...
T_5	430	426	426	...
T_6	1295	1269	1314	1289
T_7	625	622	630	...
T_8	416	416	414	...
T_9	732	725	...	727
T_{10}	427	427	...	421

period of 2651 seconds in the Gutenberg model and 2732 seconds in the Jeffreys-Bullen model. In general, the resonant frequencies for the fundamentals in the Jeffreys-Bullen model earth are lower than the frequencies of the Gutenberg model earth for the lower-order oscillations. The relations are reversed in the higher-order oscillations. The resonant frequencies of the higher overtones are virtually identical in the two models. It should be noted that no resonant frequency is listed for T_1 , since this is a degenerate case that corresponds to the rotation of the earth.

A comparison of the frequencies of oscillations of the Gutenberg and Jeffreys model earths indicates that a distinction between these two models on the basis of analysis of the frequencies of toroidal oscillations would require the identification and measurement of the resonant frequencies to 1 or 2 per cent. Since the lines are broadened by various processes (to be discussed later) it appears unlikely that a straightforward distinction between these two models is possible with present existing seismic techniques. However, the use of long-period seismometers specifically designed to record in the frequency range of a few cycles per hour coupled with modern statistical analysis might permit one to distinguish between two models as nearly alike as those of Gutenberg and Jeffreys.

The period of toroidal oscillations for the

Gutenberg and Jeffreys-Bullen model earths together with the periods obtained by *Alterman and others* [1959] and *Takeuchi* [1959] are listed in Table 4. Alterman and others studied Bullen's model B, and Takeuchi, using a variational technique, investigated a model similar to the Bullen model B. The values obtained by *Alterman and others* and those obtained in the present study are similar, again emphasizing the relative insensitivity of the periods to small variations in the distribution of elastic properties. A distinction between Bullen's model B and the Gutenberg or Jeffreys-Bullens models, on the basis of the resonant frequencies, would again require great accuracy in the determination of the frequencies. The resonant frequencies obtained by Takeuchi are significantly higher than those obtained in the present study or those obtained by Alterman and others. Since Takeuchi used a variational method in which the approximation approaches the correct solution from the high-frequency side, the results are not unexpected.

Phase velocities of long-period Love waves. *Jeanes* [1923] showed that in the limit of large l , where l is the order of the free oscillations, the free oscillations could be regarded as dispersive surface waves. For large l the phase velocities associated with the surface wave of frequency f is given by $2\pi af/(l + \frac{1}{2})$. Since the toroidal oscillations involve motion parallel to concentric spherical surfaces, they correspond to Love waves. In the limit of large l the phase velocity of Love waves approaches the shear wave velocity at the surface of the earth. Though the concept of phase velocity is hardly applicable to the low-frequency oscillations discussed in this paper, the limiting behavior of the phase velocity of Love waves at low frequencies can be obtained from the results given in Table 3. The phase velocities for the Gutenberg and the Jeffreys-Bullen model earths corresponding to the fundamental and various overtones are listed in Table 5. For the fundamental oscillations the phase velocities show a maximum at a period of 1300 sec, whereas in the Gutenberg model earth the phase velocity is 6.87 km/sec and in the Jeffreys-Bullen model earth it is 7.05 km/sec. At shorter periods the phase velocities in both the Gutenberg and Jeffreys-Bullen model earths decrease monotonically so

TABLE 5. Phase Velocities of Long-Period Love Waves

<i>l</i>	Gutenberg				Jeffreys-Bullen			
	Funda- mental <i>V</i> , km/sec	First overtone <i>V</i> , km/sec	Second overtone <i>V</i> , km/sec	Third overtone <i>V</i> , km/sec	Funda- mental <i>V</i> , km/sec	First overtone <i>V</i> , km/sec	Second overtone <i>V</i> , km/sec	Third overtone <i>V</i> , km/sec
1	...	33.3	59.1	86.5	...	33.1	60.2	86.8
2	6.04	21.3	36.3	52.3	5.84	21.4	36.3	52.6
3	6.75	16.6	26.6	37.9	6.67	16.7	26.7	37.9
4	6.87	14.2	21.4	30.0	7.05	14.3	21.3	30.1
5	6.79	12.8	18.3	25.0	6.83	12.6	18.2	25.1
6	6.70	12.2	16.2	21.8	6.71	11.9	16.3	21.9
7	6.57	11.3	14.9	19.4	6.56	11.2	14.9	19.5
8	6.43	10.7	13.9	17.7	6.50	10.8	13.8	17.7
9	6.31	10.4	13.1	16.3	6.40	10.4	13.1	16.4
10	6.18	10.0	12.6	15.3	6.28	9.9	12.6	15.4
11	6.09	9.75	12.1	14.5	6.17	9.7	12.1	14.6
12	5.98	9.47	11.7	13.9	6.06	9.5	11.7	13.9
13	5.89	9.24	11.4	13.3	5.95	9.2	11.4	13.4
14	5.80	9.02	11.1	13.0	5.83	9.1	11.2	12.9
15	5.73	8.82	10.9	12.6	5.70	8.8	10.9	12.7
16	5.68	8.69	10.6	12.3	5.60	8.8	10.7	12.4
17	5.59	8.50	10.4	12.0	5.50	8.6	10.5	12.0
18	5.52	8.34	10.2	11.8	5.40	8.4	10.3	11.9

that at a period of 400 sec (*l* = 18), the phase velocity is 5.52 km/sec in the Gutenberg model earth and 5.40 km/sec in the Jeffreys-Bullen model earth (see Fig. 5 and 6). The values obtained are not in disagreement with experimental determinations of phase velocity of Love waves and G waves [Sato, 1958].

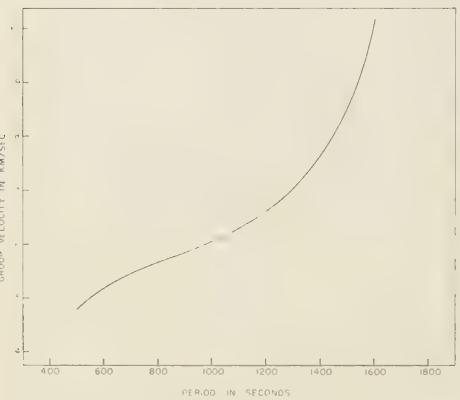


Fig. 5. Phase velocity dispersion curve for Jeffreys-Bullen model.

velocities listed in Table 5 and are shown in Figures 7 and 8. The dispersion curves show no flattening but a continual decrease in group velocity from periods of the order of 2500 sec to 400 sec. The group velocity at 400 sec is 4.4 km/sec. This corresponds to a wavelength of 2000 km. Sato [1958], in his study of the New Guinea earthquake, obtained a group velocity for the G wave of 4.5 km/sec at a wavelength of 1500 km.

Also listed in Table 5 are the phase velocities corresponding to the overtone oscillations. The phase velocities are much greater than those of the fundamental.

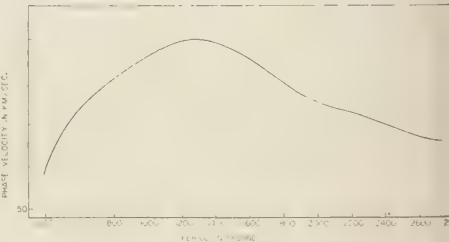


Fig. 6. Phase velocity dispersion curve for Gutenberg model.

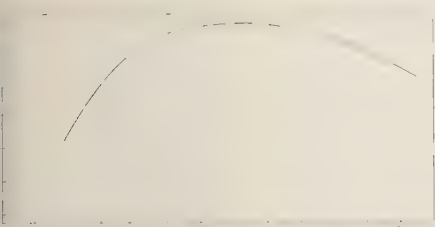


Fig. 7. Group velocity dispersion curve for Jeffreys-Bullen model.

Relative amplitudes of toroidal oscillations. The resonant frequencies of the earth are independent of the nature of the source and are fixed solely by the distribution of elastic parameters through the earth. The relative amplitudes of the toroidal oscillations depend on the depth within the earth at which the stresses are released and on the nature of the stresses. In this section we examine the relative amplitudes of toroidal oscillations excited by two types of sources. In both cases the excitation stresses are assumed to be delta functions in space and time. The extension to finite sources, having an extended time history, is immediate.

A torque excites only the toroidal oscillations. A couple excites both the toroidal and spheroidal oscillations. In the present paper we do not consider the spheroidal oscillations that are excited by a unit couple. The generation of an earthquake by breaking along a fault is more nearly approximated by a distribution of couples than by a torque. The calculations for the torque are viewed primarily in order to illustrate the difference between a source exciting only one set of oscillations and a source exciting both spheroidal and toroidal oscillations. If the

source excites only toroidal oscillations, the relative amplitudes of the oscillations will increase with mode number. If, on the other hand, the source excites both toroidal and spheroidal oscillations, the relative amplitude of the oscillation will decrease with increasing mode number. A source exciting both spheroidal and toroidal oscillations will, in general, result in a spectrum in which the amplitudes of the lines at low frequencies are greater than the amplitudes at higher frequencies. If the oscillations excited are purely toroidal, then the amplitude spectrum should increase towards high frequencies.

The relative amplitudes of toroidal oscillations for a Gutenberg model earth in which the source is at depths of 600 km, 250 km, 100 km, and 30 km are listed in Table 6. The amplitudes are normalized relative to the amplitude of ${}_1T_2$ oscillation, resulting from a unit source located at a 600-km depth. The absolute amplitude of the ${}_1T_2$ oscillation for a unit torque is 4.25×10^{-26} cm. In general, the amplitudes of the fundamental oscillations increase with increasing mode number, and the amplitudes of the overtones are less than the amplitudes of the fundamentals. The amplitudes of the fundamentals and of the overtones decrease with increasing depth of the source.

The relative amplitudes of toroidal oscillations for the Gutenberg and Jeffreys-Bullen model earths, if the source is assumed to be a unit couple, are listed in Tables 7 and 8. In both cases the amplitudes are normalized to the amplitude of ${}_1T_2$ for the source at a 600-km depth. In the Gutenberg model earth, the absolute displacement for the ${}_1T_2$ is 1.59×10^{-26} cm. The displacement for this same oscillation in the Jeffreys-Bullen model earth is 7.0×10^{-26} cm. The amplitudes of the fundamentals decrease with increasing mode number. The decrease is somewhat more rapid for the Gutenberg model earth than for the Jeffreys-Bullen model earth. The amplitudes of the fundamentals decrease with increasing depth of burial of the source. The effect is greater in the case of the Jeffreys-Bullen model earth than in the Gutenberg model earth. For the higher-order oscillations the difference between the amplitudes excited by a surface source and a 600-km source is almost a factor of 2. This is an expression of the well-known observation that the ampli-



Fig. 8. Group velocity dispersion curve for Gutenberg model.

TABLE 6. Relative Amplitudes of Toroidal Oscillations in Gutenberg Model Earth for Unit Torque

Oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth	Oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth
${}_2T_1$.152	.238	.265	.264	${}_1T_9$	1.625	2.218	2.440	2.447
${}_3T_1$	-.006	.136	.178	.179	${}_2T_9$.035	.463	.631	.668
${}_4T_1$	-.067	.055	.104	.106	${}_3T_9$	-.244	.321	.536	.553
					${}_4T_9$	-.373	.198	.464	.455
${}_1T_2$	1.00	1.06	1.085	1.082	${}_1T_{10}$	1.739	2.461	2.562	2.678
${}_2T_2$.201	.338	.381	.378	${}_2T_{10}$	-.090	.569	.671	.804
${}_3T_2$	-.010	.229	.290	.294	${}_3T_{10}$	-.275	.284	.516	.506
${}_4T_2$	-.112	.091	.176	.179	${}_4T_{10}$	-.439	.159	.489	.511
${}_1T_3$.964	1.066	1.112	1.100	${}_1T_{11}$	1.633	2.466	2.647	2.574
${}_2T_3$.231	.040	.046	.046	${}_2T_{11}$	-.163	.587	.852	.864
${}_3T_3$	-.034	.294	.039	.040	${}_3T_{11}$	-.314	.260	.488	.507
${}_4T_3$	-.165	.128	.250	.257	${}_4T_{11}$	-.423	.149	.458	.468
${}_1T_4$	1.141	1.302	1.342	1.346	${}_1T_{12}$	1.779	2.680	3.005	2.930
${}_2T_4$.211	.430	.507	.512	${}_2T_{12}$	-.253	.595	.905	.913
${}_3T_4$	-.050	.333	.482	.486	${}_3T_{12}$	-.374	.226	.509	.518
${}_4T_4$	-.197	.151	.293	.298	${}_4T_{12}$	-.384	.095	.411	.435
${}_1T_5$	1.285	.812	1.597	1.600	${}_1T_{13}$	1.790	3.059	3.445	3.422
${}_2T_5$.177	.248	.546	.549	${}_2T_{13}$	-.353	.589	.944	.947
${}_3T_5$	-.074	.189	.566	.594	${}_3T_{13}$	-.414	.225	.515	.544
${}_4T_5$	-.244	.079	.321	.364	${}_4T_{13}$	-.407	.058	.422	.428
${}_1T_6$	1.446	1.751	1.818	1.840	${}_1T_{14}$	2.005	3.311	3.820	3.883
${}_2T_6$.163	.472	.581	.570	${}_2T_{14}$	-.474	.641	1.040	1.092
${}_3T_6$	-.123	.351	.537	.509	${}_3T_{14}$	-.445	.206	.533	.532
${}_4T_6$	-.280	.178	.390	.407	${}_4T_{14}$	-.400	.070	.417	.460
${}_1T_7$	1.691	2.143	2.283	2.291	${}_1T_{15}$	2.187	3.723	4.029	3.966
${}_2T_7$.099	.468	.589	.589	${}_2T_{15}$	-.559	.515	1.106	1.133
${}_3T_7$	-.162	.398	.590	.588	${}_3T_{15}$	-.508	.195	.573	.583
${}_4T_7$	-.316	.222	.297	.394	${}_4T_{15}$	-.409	.064	.429	.454
${}_1T_8$	2.020	2.654	2.892	2.850	${}_1T_{16}$	2.911	5.183	5.911	6.104
${}_2T_8$.061	.468	.761	.594	${}_2T_{16}$	-.451	-.460	-.821	-.824
${}_3T_8$	-.204	.364	.563	.562	${}_3T_{16}$	-.556	.233	.752	.788
${}_4T_8$	-.334	.345	.411	.420	${}_4T_{16}$	-.389	.065	.433	.477

tudes of Love waves associated with deep-focus earthquakes are far smaller than the amplitudes of Love waves in near-surface earthquakes.

The amplitude spectra for a coupled source, buried at 600 km and at 30 km in the Jeffreys-Bullen model earth, are shown in Figures 9 and 10. The figures strikingly illustrate the richness of the spectra in the frequency range of 1.2 to 30 cycles/hour. Only half of the actual lines are shown, since, in addition to the toroidal oscillation, the lines associated with the

spheroidal oscillations must be added. The figures further emphasize the fact that the low frequency oscillations are excited to a greater extent for a source exciting both spheroidal and toroidal oscillations than are the higher-frequency oscillations.

The calculations listed in Tables 7 and 8 were made for an impulsive source. In general the ratio of the amplitudes at low frequencies to amplitudes at higher frequencies would be increased if the input spectra were richer in the

TABLE 7. Relative Amplitudes of Toroidal Oscillations in Gutenberg Model Earth for Unit Couple

oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth	Oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth
T_1	.201	.317	.353	.353	$1T_{10}$.042	.060	.062	.065
T_1	-.008	.180	.238	.238	$2T_{10}$.0022	.014	.016	.019
T_1	-.090	.073	.139	.142	$3T_{10}$	-.007	.007	.013	.012
					$4T_{10}$	-.011	.0038	.012	.014
T_2	1.00	1.050	1.082	1.082	$1T_{11}$.033	.048	.053	.052
T_2	.200	.337	.380	.377	$2T_{11}$	-.0033	.012	.017	.017
T_2	-.008	.228	.290	.293	$3T_{11}$	-.006	.005	.010	.010
T_2	-.112	.091	.170	.179	$4T_{11}$	-.009	.003	.009	.009
T_3	.214	.230	.247	.244	$1T_{12}$.030	.046	.051	.050
T_3	.051	.009	.010	.010	$2T_{12}$	-.004	.010	.016	.016
T_3	-.007	.065	.087	.088	$3T_{12}$	-.006	.004	.009	.009
T_3	-.037	.028	.055	.057	$4T_{12}$	-.007	.002	.007	.007
T_4	.152	.173	.179	.179	$1T_{13}$.026	.045	.050	.050
T_4	.028	.057	.067	.068	$2T_{13}$	-.005	.009	.014	.014
T_4	-.007	.044	.064	.065	$3T_{13}$	-.006	.003	.008	.008
T_4	-.026	.020	.039	.040	$4T_{13}$	-.006	.001	.006	.006
T_5	.114	.134	.142	.142	$1T_{14}$.025	.042	.049	.049
T_5	.016	.041	.048	.049	$2T_{14}$	-.006	.008	.013	.014
T_5	-.007	.031	.050	.047	$3T_{14}$	-.006	.003	.007	.007
T_5	-.021	.013	.028	.030	$4T_{14}$	-.005	.001	.005	.005
T_6	.092	.111	.115	.117	$1T_{15}$.024	.041	.045	.044
T_6	.010	.030	.037	.036	$2T_{15}$	-.006	.006	.012	.013
T_6	-.008	.022	.034	.032	$3T_{15}$	-.006	.002	.006	.007
T_6	-.018	.011	.025	.026	$4T_{15}$	-.005	.001	.005	.005
T_7	.080	.012	.011	.011	$1T_{16}$.029	.051	.038	.060
T_7	.005	.023	.028	.028	$2T_{16}$.004	-.005	-.008	-.008
T_7	-.008	.019	.028	.028	$3T_{16}$	-.005	.002	.007	.008
T_7	-.015	.011	.014	.018	$4T_{16}$	-.004	.006	.004	.005
T_8	.075	.010	.011	.011	$1T_{17}$.004	.009	.011	.012
T_8	.002	.017	.022	.022	$2T_{17}$	-.005	.004	.008	.008
T_8	-.008	.014	.021	.021	$3T_{17}$	-.007	.002	.006	.007
T_8	-.012	.013	.015	.016	$4T_{17}$	-.003	.000	.004	.004
T_9	.048	.066	.072	.072	$1T_{18}$.003	-.007	.007	.012
T_9	.001	.014	.019	.020	$2T_{18}$	-.005	.004	.008	.008
T_9	-.007	.009	.016	.016	$3T_{18}$	-.006	.001	.006	.006
T_9	-.011	.006	.014	.014	$4T_{18}$	-.003	.000	.003	.004

over frequencies. We should thus expect the amplitudes of the toroidal oscillations, excited by a unit couple, to increase toward the lower frequencies.

Perturbations to the line spectra. The idealized line spectra (Fig. 10) would, of course, never be observed on the earth. The line spectra

would be imbedded in a continuum corresponding to the random vibrations of the earth, set up by the random sources over and within the earth's surface. Furthermore, the lines are split and broadened by various processes. The discussion so far has been limited to a perfectly elastic, spherically stratified earth. Deviations

TABLE 8. Relative Amplitudes of Toroidal Oscillations for Jeffreys-Bullen Model Earth for Unit Couple

Oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth	Oscillation	Source at 600-km depth	Source at 250-km depth	Source at 100-km depth	Source at 30-km depth
${}_2T_1$.460	.727	.832	.788	${}_1T_{10}$.069	.100	.101	.107
${}_3T_1$.018	.442	.531	.521	${}_2T_{10}$	-.010	.027	.041	.037
${}_4T_1$	-.191	.178	.329	.327	${}_3T_{10}$	-.017	.015	.016	.027
					${}_4T_{10}$	-.021	.007	.023	.024
${}_1T_2$	1.00	1.100	1.145	1.129					
${}_2T_2$.214	.338	.363	.323	${}_1T_{11}$.098	.141	.161	.161
${}_3T_2$.000	.219	.288	.289	${}_2T_{11}$	-.014	.022	.161	.161
${}_4T_2$	-.112	.101	.172	.180	${}_3T_{11}$	-.016	.013	.026	.027
					${}_4T_{11}$	-.171	.006	.020	.019
${}_1T_3$.491	.549	.578	.584					
${}_2T_3$.118	.207	.229	.231	${}_1T_{12}$.061	.088	.100	.098
${}_3T_3$	-.0012	.153	.197	.204	${}_2T_{12}$	-.006	.021	.030	.031
${}_4T_3$	-.076	.061	.110	.115	${}_3T_{12}$	-.014	.009	.021	.022
					${}_4T_{12}$	-.015	.005	.016	.016
${}_1T_4$.389	.443	.455	.446					
${}_2T_4$.072	.139	.154	.155	${}_1T_{13}$.060	.093	.106	.104
${}_3T_4$	-.038	.076	.114	.112	${}_2T_{13}$	-.009	.016	.028	.028
${}_4T_4$	-.051	.035	.068	.072	${}_3T_{13}$	-.013	.008	.017	.017
					${}_4T_{13}$	-.013	.003	.014	.014
${}_1T_5$.307	.370	.391	.390					
${}_2T_5$.030	.095	.115	.119	${}_1T_{14}$.066	.098	.110	.111
${}_3T_5$	-.038	.076	.114	.112	${}_2T_{14}$	-.009	.017	.025	.024
${}_4T_5$	-.051	.035	.068	.072	${}_3T_{14}$	-.011	.006	.014	.014
					${}_4T_{14}$	-.011	.002	.011	.012
${}_1T_6$.231	.305	.276	.279					
${}_2T_6$.026	.074	.091	.085	${}_1T_{15}$.072	.120	.140	.141
${}_3T_6$	-.019	-.128	.078	.079	${}_2T_{15}$	-.011	.001	.023	.024
${}_4T_6$	-.023	.025	.054	.055	${}_3T_{15}$	-.010	.005	.012	.012
					${}_4T_{15}$	-.009	.001	.010	.011
${}_1T_7$.126	.162	.171	.171					
${}_2T_7$.001	.057	.075	.072	${}_1T_{16}$.011	.018	.020	.020
${}_3T_7$	-.016	.040	.058	.059	${}_2T_{16}$	-.094	.013	.021	.021
${}_4T_7$	-.032	.017	.041	.043	${}_3T_{16}$	-.010	.005	.011	.011
					${}_4T_{16}$	-.008	.0009	.009	.009
${}_1T_8$.111	.144	.155	.152					
${}_2T_8$.004	.044	.056	.055	${}_1T_{17}$	-.022	-.043	-.051	-.051
${}_3T_8$	-.001	.003	.041	.044	${}_2T_{17}$	-.013	.011	.022	.023
${}_4T_8$	-.027	.015	.039	.040	${}_3T_{17}$	-.010	.003	.010	.011
					${}_4T_{17}$	-.007	-.001	.008	.009
${}_1T_9$.104	.141	.149	.161					
${}_2T_9$	-.014	.038	.042	.044					
${}_3T_9$	-.014	.021	.033	.033					
${}_4T_9$	-.025	.013	.029	.030					

from perfect elasticity and sphericity will result in a broadening and a splitting of the lines in the spectrum. In addition, rotation of the earth and the inhomogeneous and anisotropic structure at the crust cause line splitting. In this section we briefly discuss these perturbation features.

The effect of imperfect elasticity on the line

spectrum is best described by the dimensionless quantity Q of the peak in the spectrum, where Q is defined as $Q = 2f/\Delta f$. Δf is the spacing between half-power points, and f is the frequency of the peak. In the earth Q has been estimated to be on the order of 100 for frequencies on the order of 10 cycles/hour [Knopoff and MacDonald,

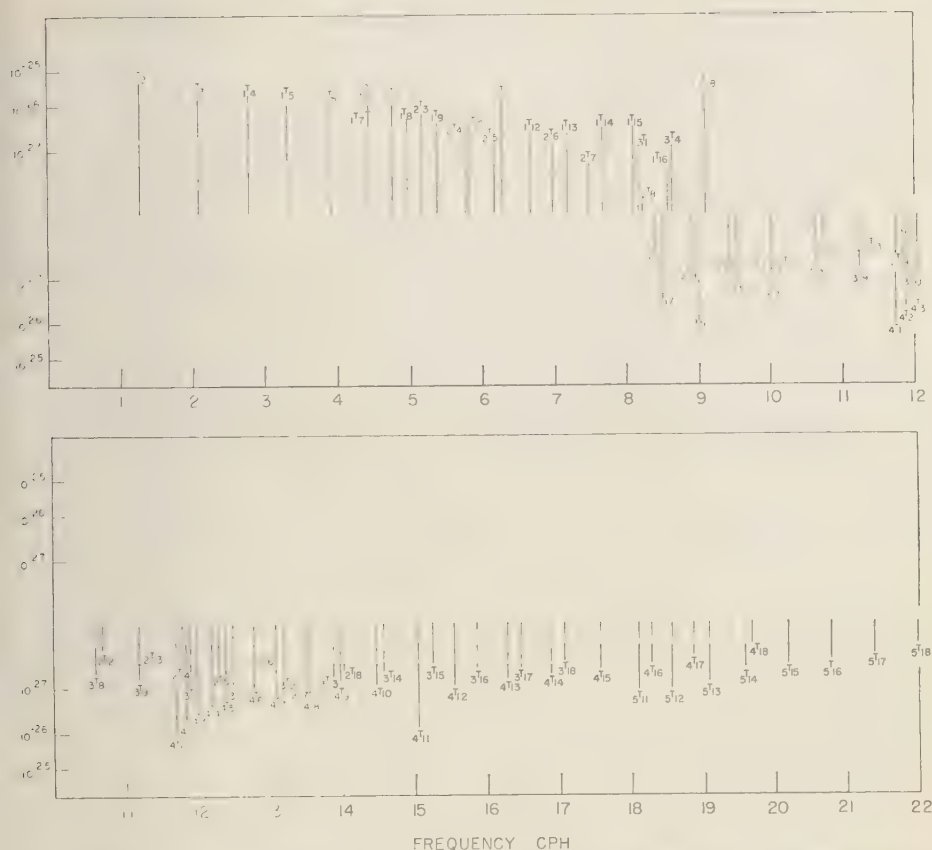


Fig. 9. Amplitude spectrum of toroidal modes for coupled source at 600-km depth; Jeffreys-Bullen model.

958]. Thus a line at 10 cycles/hour will have a width of 0.2 cycles/hour at the half-power points. Closely spaced lines, such as $1T_{15}$, $2T_1$, $3T_8$, would merge into a single asymmetric peak, regardless of resolution.

The ellipticity of the earth results in a line splitting analogous to the line splitting observed in radiation from an elliptical orbit [Sommerfeld, 1930]. The order of the splitting is of the same order of magnitude as the ellipticity. Thus this line splitting due to the ellipticity of the earth should be of the same order of magnitude as the line broadening due to the imperfection of elasticity of the earth.

In calculating the resonant frequencies of the earth we have neglected the effect of the earth's

rotation by omitting the Coriolis acceleration term in the equations of motion. A measure of the validity of neglecting the Coriolis acceleration is provided by the ratio of the frequency of rotation of the earth to the frequency of the normal mode. Since the longest mode has a period of less than an hour, the effect of rotation in shifting the resonant frequency should be small.

An additional effect of the earth's rotation and also of the inhomogeneity of the crust is the rotation of the plane of polarization of a propagating pulse, which can be thought of as a superposition of normal modes. Although the effect of the earth's rotation is presumably negligible, it is possible that the crustal effect

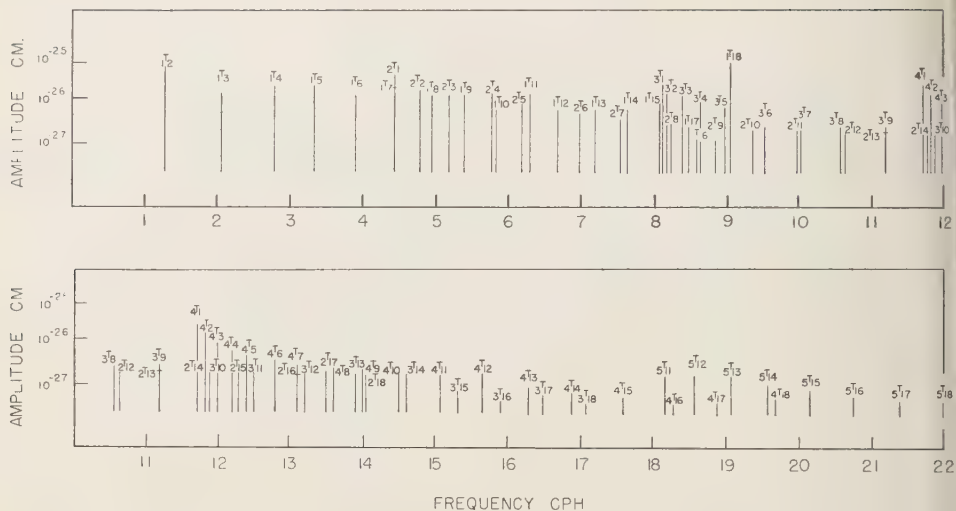


Fig. 10. Amplitude spectrum of toroidal modes for coupled source at 30-km depth; Jeffreys-Bullen model.

may be significant. The inhomogeneity of the crust has the further effect of splitting the lines and coupling the toroidal and coupled oscillations. The order of magnitude of this effect has not been investigated.

Ray theory-mode theory correspondence: an example. In order to demonstrate the relationship between the free oscillations and ray theory we examine a problem similar to the toroidal oscillation problem but simpler in mathematical detail.

Consider the problem of wave motion in a homogeneous sphere. Let the scalar potential be ϕ . A source is located at $(r_0, 0)$ in spherical coordinates, and the potential of the source is

$$\phi_0 = \frac{e^{ikR}}{R} = ik \sum_{l=0}^{\infty} (2l+1) \cdot P_l(\cos \theta) j_l(kr_<) h_l^{(1)}(kr_>) \quad (27)$$

where $R^2 = r^2 + r_0^2 - 2rr_0 \cos \theta$ and $k = \omega/c$ is the wave number and $h_l^{(1)}$ is the spherical Hankel function. Then the total potential in the sphere is

$$\phi = ik \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) \cdot [j_l(kr_<) h_l^{(1)}(kr_>) + A_l j_l(kr)] \quad (28)$$

where A_l is determined from the boundary condition at $r = a$, the radius of the sphere. For simplicity, we consider the boundary condition $\partial\phi/\partial r = 0$ at $r = a$. Using the orthogonality relations for Legendre polynomials, we find

$$A_l = -j_l(kr_0) h_l^{(1)'}(ka) / j_l'(ka) \quad (29)$$

Now let $r \rightarrow a$ so that the potential on the surface is

$$\phi = ik \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) [h_l^{(1)'}(ka) j_l'(ka) - h_l^{(1)'}(ka) j_l(ka)] j_l(kr_0) / j_l'(ka) \quad (30)$$

which can be written

$$\phi = \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) j_l(kr_0) / ka^2 j_l'(ka) \quad (31)$$

by using the Wronskian of the spherical Bessel functions.

Poles of (31) as a function of k occur at the roots of $j_l'(ka) = 0$. We call the periods corresponding to the zeros of $j_n'(ka)$ the free oscillations of the sphere. The Fourier inversion of (31) over frequency leads to the time solution

$$\psi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi e^{-i\omega t} d\omega \quad (32)$$

The inversion integral can be evaluated at the poles of $j_l'(ka)$ by Cauchy's theorem

$$\begin{aligned} &= i \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) \sum_{n=-\infty}^{\infty} \\ &\quad \cdot [c j_l(n k_i r_0) / n k_i a^3 j_l''(n k_i a)] e^{-i n \omega t} \\ &= \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) \\ &\quad \cdot \sum_{n=1}^{\infty} (2 c j_l(n k_i r_0) / n k_i a^3 j_l''(n k_i a)) \\ &\quad \cdot \sin(n \omega t) \end{aligned} \quad (33)$$

Equation (33) represents the time response, due to a delta function source, as the weighted superposition of the free oscillations of the sphere. The convergence of (33) looks questionable but can be demonstrated.

An alternative representation of the response due to a point source in a sphere is afforded by ray theory. That the ray solution can be obtained from expressions such as (31) can be shown rather simply. First we make the change of variable $k = iq$, $s = i\omega$, and (31) becomes

$$= - \sum_{l=0}^{\infty} (2l+1) P_l(\cos \theta) j_l(qr_0) / q a^2 j_l'(qa) \quad (34)$$

where j_l is the modified spherical Bessel function of the first kind. The sum in (34) can be expressed as an integral by Watson's lemma

$$\begin{aligned} &= \pi i^{-1} \int_L d\lambda \lambda \sec \lambda \pi P_{\lambda-1/2} \\ &\quad \cdot (-\cos \theta) j_{\lambda-1/2}(qr_0) / q a^2 j_{\lambda-1/2}'(qa) \end{aligned} \quad (35)$$

the path L is shown in Figure 11.

Evaluating (35) at the poles of $\sec \lambda \pi$ yields the series equation (34). The ray solution is obtained by evaluating (35) by the saddlepoint method. From previous work [Gilbert and Snopoff, 1959] we know that the saddlepoints occur for imaginary λ , $|\lambda| < qr_0$. Let $\lambda = i\mu$. Then

$$\begin{aligned} &= -\pi i \int_{c-i\infty}^{c+i\infty} d\mu \mu \operatorname{sech}(\mu \pi) P_{i\mu-1/2} \\ &\quad \cdot (-\cos \theta) j_{i\mu-1/2}(qr_0) / q a^2 j_{i\mu-1/2}'(qa) \end{aligned} \quad (36)$$

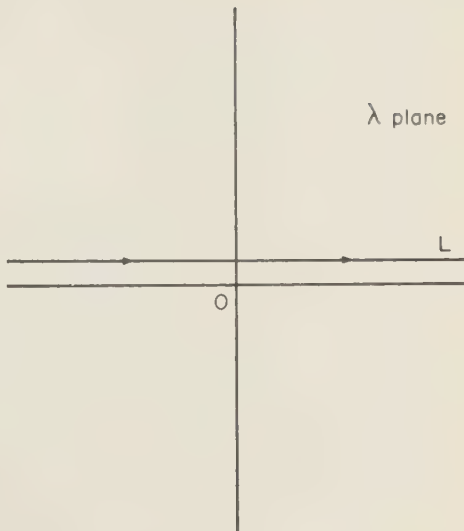


Fig. 11. Path of integration in λ plane for series transformation by Watson's lemma.

By a well-known Tauberian theorem the behavior of (36) near the wave fronts of the reflected rays is obtained by letting $s \rightarrow \infty$. Then the saddlepoints will occur for large $\mu < qr_0$, and we use the transitional asymptotic form for the Bessel functions:

$$\begin{aligned} j_{i\mu-1/2}(z) &\approx \frac{e^{\mu \pi/2}}{\partial z^{1/2} (z^2 - \mu^2)^{1/4}} \cdot [ie^{-\eta(z)} + e^{\eta(z)}] \\ j_{i\mu-1/2}'(z) &\approx -\frac{(z^2 - \mu^2)^{1/4} e^{\mu \pi/2}}{\partial z^{3/2}} \cdot [ie^{-\eta(z)} - e^{\eta(z)}] \end{aligned} \quad (37)$$

where

$$\eta(z) = (z^2 - \mu^2)^{1/2} - \cos^{-1}(\mu/z).$$

From (37) we find

$$\begin{aligned} &j_{i\mu-1/2}(z) / j_{i\mu-1/2}'(x) \\ &\approx \frac{x^{3/2}}{z^{1/2} (z^2 - \mu^2)^{1/4} (x^2 - \mu^2)^{1/4}} \\ &\quad \cdot \left[\frac{e^{\eta(z)} + ie^{-\eta(z)}}{e^{\eta(x)} - ie^{-\eta(x)}} \right] \\ &= \frac{x^{3/2}}{z^{1/2} (z^2 - \mu^2)^{1/4} (x^2 - \mu^2)^{1/4}} \end{aligned}$$

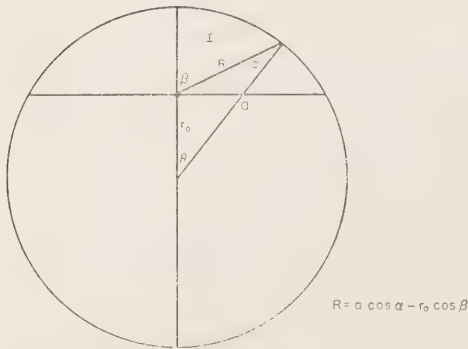


Fig. 12. Geometrical interpretation of the saddlepoint for direct ray from source to surface

$$\begin{aligned} & \cdot [e^{\eta(z) - \eta(x)} + ie^{-\eta(z) - \eta(x)}] \\ & \cdot \sum_{j=0}^{\infty} [ie^{-2\eta(x)}]^j. \end{aligned} \quad (38)$$

The expansion in (38) is suggestive of a ray interpretation. To derive the ray solution from (36) we use the asymptotic expansion of $\text{sech } \mu\pi$ and $P_{i\mu-1/2}(-\cos \theta)$:

$$\begin{aligned} \text{sech } \mu\pi &= 2 \sum_{m=0}^{\infty} e^{-\mu(2m+1)\pi + m\pi i} \\ P_{i\mu-1/2}(-\cos \theta) &\approx (2\pi\mu \sin \theta)^{-1/2} \\ &\cdot [e^{\mu(\theta-\pi)} + e^{\mu(\pi-\theta)}] \quad 0 < \theta < \pi \end{aligned} \quad (39)$$

Using (38) and (39) in (36), we have

$$\begin{aligned} \phi &= 2\pi i^{-1} \int_{c-i\infty}^{c+i\infty} d\mu \mu \sum_{m=0}^{\infty} e^{-\mu(2m+1)\pi + m\pi i} \\ &\cdot [e^{\mu(\pi-\theta)} + e^{\mu(\theta-\pi)}] (2\pi\mu \sin \theta)^{-1/2} \\ &\cdot \frac{[e^{\eta(qr_0) - \eta(qa)} + ie^{-\eta(qr_0) - \eta(qa)}]}{(ar)^{1/2} (q^2 r_0^2 - \mu^2)^{1/4} (q^2 a^2 - \mu^2)^{1/4}} \\ &\cdot \sum_{j=0}^{\infty} [ie^{-2\eta(qa)}]^j \end{aligned} \quad (40)$$

The evaluation of (40) by the saddlepoint method yields the ray solution. Consider, for example, the term

$$\phi_0 = 2\pi i^{-1} \int_{c-i\infty}^{c+i\infty} d\mu \mu \frac{e^{-\mu\theta + \eta(qr_0) - \eta(qa)}}{(2\pi\mu ar \sin \theta)^{1/2} (q^2 r_0^2 - \mu^2)^{1/4} (q^2 a^2 - \mu^2)^{1/4}} \quad (41)$$

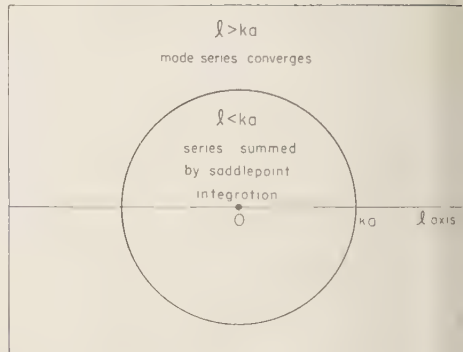


Fig. 13. Region of convergence of mode series and region of effect of saddlepoint integrations.

The saddlepoint in (41) occurs for $-\theta - \cos^{-1}(\mu_0/qr_0) + \cos^{-1}(\mu_0/qa) = 0$.

Let $\mu_0 = qr_0 \sin \beta = qa \sin \alpha$. The geometrical interpretation of the saddlepoint is shown in Figure 12.

The saddlepoint exists only in the region I and (41) represents the direct ray from the source to the surface in region I . The exponent evaluated at μ_0 is $-qR$, and the integral in (41) becomes

$$\phi_0 = \frac{2e^{-qR}}{R}, \quad \psi_0 = \frac{2\delta(t - R/c)}{R} \quad (42)$$

The factor 2 represents the well-known doubling of the potential on the surface. The three remaining terms in (40) for $m = 0$, $j = 0$ represent the direct ray in the remaining three sections of the sphere accordingly as the points on the surface are above or below the source and $\theta \geq 0$. The terms for $m \neq 0$, $j \neq 0$ represent the multiply reflected rays in the sphere and can be also evaluated by the saddlepoint method. The evaluation will not be carried out, as our purpose is to show the relationship between the modes (free oscillations) and ray theory. Jeffreys and Lapwood [1957] have considered the multiple-reflection problem.

Thus the response due to a source in a sphere can be regarded as the superposition of the free oscillations of the sphere excited by the source

as the superposition of the direct and internally reflected rays.

Since the original series equations (30) and (31) begin to converge for $l > ka$, the effect of the saddlepoint integration is to sum those terms $l < ka$ in the series. The sum of rays is then expressed as the sum of modes. Each saddlepoint integral selects the mode contributions which correspond to its ray event (Fig. 13).

Any seismogram may be analyzed as the superposition of many ray events and surface arrivals or as the superposition of free oscillations. The former approach has been used extensively and has led to a considerable understanding of the structure of the earth. The latter approach has not been investigated in detail.

REFERENCES

- Ben-Menahem, Z., H. Jarosch, and C. L. Pekeris, Oscillations of the earth, *Proc. Roy. Soc. London A*, **252**, 80-95, 1959.
- Bioff, H., J. C. Harrison, W. H. Munk, and J. B. Slichter, Searching for the earth's free oscillations, *J. Geophys. Research*, **64**, 1334-1337, 1959.
- Bioff, H., Long period waves; Progress report, Seismol. Lab., Calif. Inst. Technol., *Trans. Am. Geophys. Union*, **35**, 984-985, 1954.
- Ben-Menahem, A., Free non-radial oscillations of the earth, *Geofis. pura e appl.*, **43**, 23-35, 1959.
- Van der Laan, E. C., The density within the earth, *Verhandel. Ned. Geol.-Mijnbouwk. Genoot.*, **Geol. Ser.**, **18**, 23-41, 1957.
- Allen, K. E., *An Introduction to the Theory of Seismology*, 2nd Ed., Cambridge University Press, 1953.
- Ben-Menahem, J., Theory and computation of properties of surface waves on layered media (abstr.), *J. Geophys. Research*, **64**, 1101, 1959.
- Jobert, F., and L. Knopoff, Scattering of impulsive elastic waves by a rigid cylinder, *J. Acoust. Soc. Am.*, 1959 (in press).
- Jobert, F., and G. J. F. MacDonald, Free oscillations of the earth (abstr.), *J. Geophys. Research*, **64**, 1103, 1959.
- Ben-Menahem, B., Wave velocities at depth between 40 and 600 kilometers, *Bull. Seis. Soc. Am.*, **43**, 223-232, 1953.
- Ben-Menahem, N. A., The dispersion of surface waves on multilayered solid media, *Bull. Seis. Soc. Am.*, **43**, 17-34, 1953.
- Christoff, P., Ueber die elastischen Schwingungen einer isotropen Kugel, *Crelle*, **88**, 131-145, 1880.
- Jeans, J. H., On the vibrations and stability of a gravitating planet, *Phil. Trans. Roy. Soc. London, Ser. A*, **201**, 157-184, 1903.
- Jeans, J. H., The propagation of earthquake waves, *Proc. Roy. Soc. London A*, **102**, 554-574, 1923.
- Jeffreys, H., *The Earth*, 4th Ed., Cambridge University Press, 1959.
- Jeffreys, H., and E. R. Lapwood, The reflexion of a pulse within a sphere, *Proc. Roy. Soc. London A*, **241**, 455-479, 1957.
- Jobert, N., Evaluation de la periode d'oscillation d'une sphere elastique heterogene, par application du principe de Rayleigh, *Compt. rend.*, **243**, 1230-1232, 1956.
- Jobert, N., Sur la periode propre des oscillations spheriodales de la Terre, *Compt. rend.*, **244**, 921-922; **245**, 1941-1943, 1957.
- Knopoff, L., and G. J. F. MacDonald, Attenuation of small amplitude stress waves in solids, *Revs. Modern Phys.*, **30**, 1178-1192, 1958.
- Lamb, H., On the vibrations of an elastic sphere, *Proc. London Math. Soc.*, **13**, 189-212, 1882.
- Love, A. E. H., *Some Problems of Geodynamics*, Cambridge University Press, 1926.
- Matumoto, T., and Y. Sato, On the vibrations of an elastic globe with one layer. The vibration of the first class, *Bull. Earthquake Research Inst., Tokyo Univ.*, **32**, 247-258, 1954.
- Molodenski, M. S., Elastic tides, free nutation, and some problems of the earth's structure, *Trudy Geofiz. Inst., Akad. Nauk SSSR Sbornik Statei*, **19**, 1953.
- Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill Book Co., New York, 1953.
- Pekeris, C. L., and H. Jarosch, The free oscillations of the earth; in *Contributions in Geophysics*, Pergamon Press, Los Angeles, 1958.
- Lord Rayleigh, On the dilatational stability of the earth, *Proc. Roy. Soc. London A*, **77**, 486-499, 1906.
- Sato, Y., Attenuation, dispersion, and the wave guide of the G-wave, *Bull. Seis. Soc. Am.*, **48**, 231-252, 1958.
- Sommerfeld, A., *Wave-mechanics*, Methuen & Co., London, p. 112, 1930.
- Stoneley, R., The elastic yielding of the earth, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, **1**, 356-359, 1926.
- Takeuchi, H., On the earth tide of the compressible earth of variable density and elasticity, *Trans. Am. Geophys. Union*, **31**, 651-689, 1950.
- Takeuchi, H., Torsional oscillations of the earth and some related problems, *Geophys. J., Roy. Astron. Soc.*, **2**, 89-100, 1959.
- Thomson, W. T., Transmission of elastic waves through a stratified solid medium, *J. Appl. Phys.*, **21**, 89-93, 1950.

(Manuscript received October 21, 1959.)

Studies in the Theory of Shock Propagation in Solids

WILLIAM BAND

*Department of Physics
Washington State University, Pullman, Washington
and
Poulter Laboratories, Stanford Research Institute
Menlo Park, California*

***Abstract.** In Part I a single-parameter visco-elastic model of a shear-yielding solid is defined for which a permanent-regime solution of the equations of motion exists for any finite compression. The 'profile' of this compression as a function of distance is obtained in the form of an integral which can be evaluated when the velocity of propagation is known as a function of final compression. It is assumed that the permanent-regime solution approximates actual shock waves, and that the velocity of the permanent-regime profile equals the shock velocity. Observed shock speeds are used to compute shock profiles in a number of metals. The maximum slope of the profile for any one metal increases with increasing compression. The limiting value of the maximum slope as the volume is extrapolated to zero gives a numerical estimate of the viscosity parameter, and this has been done for Al, Pb, Sn, Zn, and Zr.

In Part II Zener's linear theory of anelasticity has been generalized to materials with cubic crystal structure. The theory of the propagation and attenuation of plane waves, both longitudinal and transverse, along a principal axis of the crystal is presented. The combined effects of relaxation mechanisms and thermal diffusion are included. The significance of the results for the theory of shock propagation are discussed and several questions are raised for later discussion.

In Part III the general equations for propagation of steady-state compression profiles in shear-yielding, heat-conducting, anelastic solids are given. Methods of solution by successive approximation are developed for Hookean solids, with both adiabatic and isothermal (very steep) profiles, and for non-Hookean solids with shock profiles. The results of Part I are corrected to include the effects of thermal conductivity and anelasticity. Heating aftereffects of shocks are discussed, including the effects of irreversible heating due to viscous yielding, etc., and it is shown how the temperature of the solid after passage of a shock profile may be calculated.

Part I. The Permanent-Regime Solution

INTRODUCTION

Shock propagation velocities in solids have in the past been successfully described in terms of a hydrodynamic model in which the rigidity is assumed to be effectively zero [Rice and others, 1958]. No detailed theoretical discussion of this model has been given, but it is generally justified by the remark that shear yielding could be complete under the high stresses present in a shock. There remains some experimental uncertainty whether the strictly hydrodynamic model should be modified to conform with the von Mises, or some similar, criterion for plastic yielding [Jaeger, 1956] of a linearly compressed solid. A decision on this question depends on a precise comparison of the isotherms deduced from shock data [Walsh and others, 1957] with those from static uniform compression

[P. W. Bridgman; see Walsh and others, 1957, for details]. The major objection to the model is that shear yielding normally takes a time that is long compared with the time of duration of a shock front. However, the effectiveness of the generation of dislocations by high stresses in facilitating shear yielding has been noted recently [Dwall and Koehler, 1959], and this approach promises to clear up this objection.

A detailed discussion of the shear-yielding process, together with all other important irreversible effects, such as thermal diffusion, phonon scattering, and viscosity, should lead eventually to a definite shock profile, depending essentially on the relaxation times for the several processes. Such a discussion has already been given for shocks in an ideal gas [Hirschfelder and others, 1954], but the solid

presents a more difficult problem both because of the more complicated (even if known) equation of state and because of the shear-yielding process.

As a first approach to such a discussion we give here a 'permanent-regime solution' [Rayleigh, 1910] for the shock profile in terms of a single parameter which, for lack of a better name, we call the shear-yielding viscosity. The physical model is as follows. Let an infinite slab of material be subject to a normal stress across its parallel faces. Under zero stress it has density ρ_0 . If the negative stress is increased at a sufficiently slow rate, the linear compression is an elastic single-valued function of the normal negative stress component p^* , up to a critical negative stress p_s^* and critical volume ratio $y_s = \rho_0/\rho_s$; but as p^* increases beyond p_s^* , the compression 'discontinuously' increases because of the yielding process. We can leave open the exact extent of this yielding, so that the theory applies equally well whether the hydrostatic or the plastic yield model is accepted.

If the negative stress is increased at a finite rate, there is no change in the critical stress, and the yielding still occurs at the same point; but with yielding the material acquires viscosity, so that a higher negative stress is required to cause the compression to occur at a finite rate. The relation between stress and strain for rapidly increasing negative stress is then of the form

$$p^*(y) = K(y) - G(y) dy/dt \quad (1)$$

where $G(y)$ is the shear-yielding viscosity as a function of the ratio y , and $K(y)$ is the negative stress component corresponding to slow linear compression. In our model $K(y)$, as a function of y , has a discontinuous first derivative; it contains the full effects of rigidity when $p^* < p_s^*$, but takes on the 'yielded value' when $p^* > p_s^*$. It is not necessary to specify its dependence upon y further than this. The viscosity $G(y)$ is also a 'discontinuous' function, being practically zero for $y > y_s$ and nonzero for $y < y_s$; its exact dependence upon y in the latter range need not be specified until specific cases are examined. Equation (1) indicates that for $y < y_s$ and dy/dt negative, the negative stress p^* is greater than the yield value by the viscous term. The dimensionality of $G(y)$ is stress multiplied by time;

it represents the relaxation characteristics of the material and controls the shock profile.

Note that we do not need to specify the behavior of the material under decreasing negative stress in order to discuss the solution of 'permanent regime,' so questions of permanent deformation and hysteresis do not arise and need not be specified in the model.

Figure 1 illustrates equation (1). The full line is $K(y)$, and the dotted line is $p^*(y)$. The fact that this is a straight line is proved later in the paper (see equation 22).

THE EQUATION OF MOTION

Let v be the speed of the material in the direction (normal to the slab face); then the equation of motion of a given mass element ρdx with unit area section is

$$(d/dt)(\rho v dx) = -dx \partial p^*/\partial x \quad (2)$$

where d/dt is the convective derivative. Since the mass of the element is a constant of the motion, this equation simplifies to

$$\rho dv/dt = -\partial p^*/\partial x \quad (3)$$

or

$$\rho \partial v/\partial t + \rho v \partial v/\partial x = -\partial p^*/\partial x$$

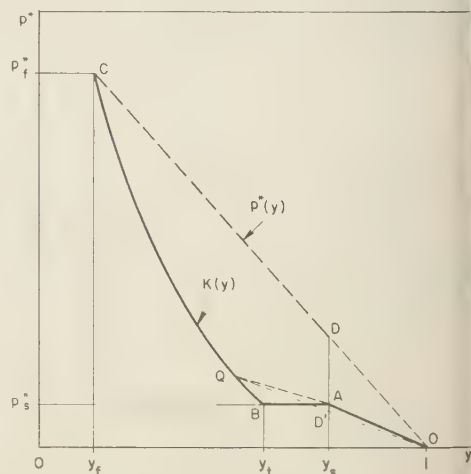


Fig. 1. Longitudinal stress versus compression. BC is the Hugoniot curve of final stress versus final compression; DC is the stress in the shock profile corresponding to the final compression y_f at the point C ; AB is the yield; and OA is the elastic curve.

Using the equation of mass conservation, we can write

$$\partial(\rho v)/\partial t + \partial(\rho v^2)/\partial x = -\partial p^*/\partial x \quad (4)$$

Noting that in (1) dy/dt is the convective derivative (referring to a given element of material), we have for the equation of motion

$$\begin{aligned} \partial(\rho v)/\partial t + \partial(\rho v^2)/\partial x &= -(\partial/\partial x)p^*(y) \\ &= -(\partial/\partial x)\{K(y) - G(y)(\partial y/\partial x) \\ &\quad + v\partial y/\partial x\} \end{aligned} \quad (5)$$

We take a coordinate system in which the undisturbed material is moving with speed U_f in the x direction and seek a value of U_f such that a solution of (5) exists with all partial derivatives equal to zero:

$$\partial(\rho v)\{ \rho v^2 + K(y) - G(y)v \partial y/\partial x \} = 0 \quad (6)$$

The 'front' will be taken at $x = 0$; to the left, $x < 0$, the density ρ_0 is uniform, the speed is uniform ($v = U_f$), and $y = 1$; to the right $x > 0$, $y < 1$, and the speed is reduced by the particle velocity u ;

$$v = U_f - u$$

$$x(y) = \int_{y'=1}^{y'=y} \frac{U_f G(y') y' dy'}{\rho_0 U_f^2 (y' - y_s) + K(y') - K(y_s) + U_f G(y_s) y_s (\partial y/\partial x)_s} \quad (12)$$

where u is a function of y . Conservation of mass requires that

$$(U_f - u)\rho = U_f \rho_0$$

where ρ is the density (function of y) in the profile; thus

$$v = (U_f - u) = y U_f, \quad y < 1 \quad (7)$$

The equation of motion, (6), is then a differential equation for y as a function of x :

$$\begin{aligned} \partial(\rho v)\{ \rho_0 y U_f^2 + K(y) \\ - G(y) y U_f \partial y/\partial x \} = 0 \end{aligned} \quad (8)$$

We assume that as x increases, the value of y asymptotically approaches a final value y_f . Thus

$$x \rightarrow \infty, \quad y \rightarrow y_f, \quad \partial y/\partial x \rightarrow 0$$

The first integral of (8) then reads

$$\begin{aligned} U_f G(y) y \partial y/\partial x &= \rho_0 U_f^2 (y - y_f) \\ &\quad + K(y) - K(y_f) \end{aligned} \quad (9)$$

The variables separate, and we have at once the profile in the form of the position x as a function of the volume ratio y :

$$x(y) = \int_{y'=y_s}^{y'=y} \frac{U_f G(y') y' dy'}{\rho_0 U_f^2 (y' - y_f) + K(y') - K(y_f)} \quad (10)$$

x is taken zero at $y = y_s$ in describing the shock profile.

The lower limit of this integral has been set at y_s , because (2) is reliable only where the various functions are well behaved. We have assumed that $K(y)$ has a discontinuous derivative at y_s , so that $\partial p^*/\partial x$ would have an infinity there. The integral in (10) extends only over a domain in which (2) is valid, and the jump conditions at the discontinuity will be considered separately.

We can, of course, carry out a similar integration for the range $1 \geq y > y_s$. Replacing (9), we now have

$$\begin{aligned} U_f G(y) y \partial y/\partial x &= \rho_0 U_f^2 (y - y_s) + K(y) \\ &\quad - K(y_s) + U_f G(y_s) y_s (\partial y/\partial x)_s \end{aligned} \quad (11)$$

and, in place of (10),

$$x(y) = \int_{y'=1}^{y'=y} \frac{U_f G(y') y' dy'}{\rho_0 U_f^2 (y' - y_s) + K(y') - K(y_s) + U_f G(y_s) y_s (\partial y/\partial x)_s} \quad (12)$$

We recall that $G(y)$ is, on the present model, almost zero throughout this range of compression values. If $G(y)$ were mathematically zero, two terms would disappear from (11) and the step to (12) could not be taken. But we adopt a more physical point of view and regard $G(y)$ as very small, and $\partial y/\partial x$ attains very large values. Equation (12) is then valid, but now, because $G(y)$ is very small, we conclude that $x(y)$ is also very small. In other words, the density increases almost discontinuously from ρ_0 to ρ_s at $x = 0$. In Figure 1 the points 0 and D are physically coincident, and any line joining them would be meaningless. As is mentioned above, because K and G are 'discontinuous' at $x = 0$ we cannot use the differential equations to determine the way in which the stress profiles for the ranges $1 \geq y > y_s$ and $y_s > y \geq y_f$ join together; for this we have to resort to the standard jump-condition technique.

THE HUGONIOT JUMP CONDITIONS

The profile integral in (10) can be evaluated

when the functions $G(y)$ and $K(y)$ are given. Such a calculation would yield a value for the gradient $\partial y/\partial x$ as a function of x . From (1) we could then obtain the longitudinal stress as a function of x in this profile. In particular, by using $(\partial y/\partial x)_{y=y_s=0}$ or $(\partial y/\partial x)_{x=+0}$ in (1) we have for the negative longitudinal stress at the beginning of the profile

$$p_0^* = K(y_s) - U_f G(y_s) y_s (\partial y/\partial x)_{x=+0} \quad (13)$$

But from (9) we also have

$$U_f G(y_s) y_s (\partial y/\partial x)_{x=+0} = \rho_0 U_f^2 (y_s - y_f) + K(y_s) - K(y_f) \quad (14)$$

and so (13) becomes

$$U_f^2 = \frac{K(y_f) - p_0^*}{\rho_0 (y_s - y_f)} = \frac{\rho_f \rho_s [K(y_f) - p_0^*]}{\rho_0^2 (\rho_f - \rho_s)} \quad (15)$$

Independently of this, conservation of mass and momentum across the plane $x = 0$ between zero pressure and the negative stress p_0^* yield the Hugoniot jump condition

$$U_f^2 = \rho_s p_0^* / [\rho_0 (\rho_s - \rho_0)] \quad (16)$$

Note that here U_f is the same speed as that which appears in (15); it is *not* the speed of a shock producing a final density ρ_s , and p_0^* is *not* the stress required to produce such a final density.

Eliminating U_f^2 between (15) and (16), we find

$$p_0^* = K(y_f)(1 - y_s)/(1 - y_f) \quad (17)$$

and (16) becomes

$$U_f^2 = (\rho_f/\rho_0) K(y_f)/(\rho_f - \rho_0)$$

Because at y_f the rate term vanishes, $K(y_f)$ is just p_f^* , and our result reads

$$U_f^2 = p_f^* / [\rho_0 (1 - y_f)] \quad (18)$$

which we recognize as the standard Hugoniot relation [Cole, 1948] for the complete shock profile; this equation is generally employed in the hydrodynamical model where the structure of the profile is ignored.

The situation is illustrated in the diagram showing $K(y)$ versus y (Fig. 1). The full curve $OABC$ would represent the behavior of the material if the viscosity of the shear yielding

process could be suppressed. The segment OA is the elastic compression with rigidity full operative. The curve BC is the Hugoniot curve of the hydrodynamical model. The broken line DC is the total negative stress component p occurring in the shock, including the part required to overcome the shear-yielding viscosity. The jump AD is the difference between p_f^* given by (17), and the stress p_s^* required to initiate the yielding in a static compression— p_s^* represents the extra stress required to initiate the finite rate of compression against the viscosity.

THE PRESSURE PROFILE IN THE SHOCK

For the permanent-regime solution equation (1) reads

$$p^*(y) = K(y) - G(y) U_f y \partial y/\partial x \quad (19)$$

From (9) this can be written

$$p^*(y) = K(y_f) - \rho_0 U_f^2 (y - y_f) \quad (20)$$

Then from (18) we have, eliminating $K(y_f) = p_f^*$

$$p^*(y) = \rho_0 U_f^2 (1 - y) \quad (21)$$

valid for $y < y_s$.

An alternative form of the same result is

$$p^*(y) = p_f^* (1 - y)/(1 - y_f) \quad (22)$$

The graphical significance of this result is that in Figure 1, the curve DC is necessarily a straight line, which, when projected, must pass through the origin. Note that (17) has already indicated that the points O , D , and C lie on a straight line. No matter how the compression depends on position through the function $y(x)$, the negative stress p^* will also depend on position in such a way as to maintain this linear relation between stress and volume ratio. Moreover, the result is completely independent of the equation of state of the material.

From the physical meaning of the negative stress difference AD , it must be positive. We shall return later to discuss shocks so weak that, for example, the maximum occurs at Q in Figure 1 and the corresponding D' falls below A .

THE DENSITY PROFILE IN THE SHOCK

We now consider only shocks so strong that we can be sure that p_0^* given by (17) is greater

in p_* . Then (18) is a general relation between shock velocity and final compression. We wish to consider shocks of various final compressions, and find it convenient to modify the notation slightly. Thus, if the final compression were given by the volume ratio y' , the corresponding shock velocity would be U' , where

$$U'^2 = K(y')/[\rho_0(1 - y')] \quad (23)$$

It follows from (18) if we note that p_* is identical with $K(y_f)$, the viscous term being zero at the final compression (zero rate). We may now use this relation to modify (10) thus:

$$y) = \int_{y'=y_s}^{y'=y} \frac{U_f G(y') y' dy'}{\rho_0(1 - y')(U'^2 - U_f^2)} \quad (24)$$

where U' signifies the velocity a shock would have if its final volume ratio were y' .

Experimental data on shock and particle velocities actually provide us with values of y' for fairly wide ranges of y' values for a considerable number of solids. For a first examination of the data, $G(y')$ will be taken as a constant—which essentially means that it will be represented by a mean value. We then compute the function

$$\begin{aligned} y) &= x \rho_0 / G \\ &= \int_{y'=y_s}^{y'=y} \frac{U_f y' dy'}{(1 - y')[U(y')^2 - U_f^2]} \quad (25) \end{aligned}$$

Experimental data [Walsh and others, 1957] give the shock velocity $U(y')$ for the final compression y' . Linear interpolations between sufficiently closely spaced points make the integral easy to evaluate. The major uncertainty in the result is in the value to be used for y_s in each case. Surprisingly enough, the onset of the profile is quite slow, and variations in y_s make significant changes in the shape of the profiles. An arbitrary value, namely $y_s = 0.999$, was accepted for all the profiles computed; smaller values would have no other effect than to shorten the 'tail' of the curve at the front.

The first set of diagrams, Figures 2 to 4, gives the profiles for shocks in a number of solids taken to the maximum final compressions given in the data of Walsh and others [1957]. Figures 5 and 6, show profiles for Pb and Al, taken to a number of different final compressions. Figures 7 and 8 show profiles computed from recent

data published by Al'tshuler and others [1958]; some of these data are also included in Figures 5 and 6.

It is worth emphasizing here that these profiles are permanent-regime solutions and are not to be identified with actual profiles as produced in the experimental arrangements used to obtain the equation of state. One may regard the real profiles as approximating permanent-regime profiles, and tending, by dissipative mechanisms, towards such profiles with time.

It is interesting to note the dependence of maximum slope of the shock profile with respect to the maximum compression. If S is the 'slope' of the profile curves at their points of inflexion, i.e., the maximum rate of change of the volume ratio with respect to distance, the dependence of S on y_f has been fitted to the following equation:

$$y_s - y_f = \left(\frac{S - S_0}{S_m - S_0} \right)^v \quad (29)$$

The constant S_0 may be interpreted as the slope of the profile of a weak shock, just sufficient to initiate the shear-yielding process, and S_m as the ideal maximum possible slope of the profile for the ideal maximum possible compression (zero volume). Table I gives the parameters found from a log-plot of the data. Because the units of velocity used here are millimeters per microsecond, the expression for the slope in terms of viscosity measured in cgs units is

$$S = 10^{-5} (G/\rho_0) \partial(\rho_0/\rho)/\partial x \quad (30)$$

If we consider the maximum space rate of compression to be given by the ratio of the maximum change in volume ratio to the minimum distance and take the latter as the lattice spacing for an order of magnitude estimate, we have

$$\text{maximum of } \partial(\rho_0/\rho)/\partial x_{\text{inflex}} = 1/D \quad (31)$$

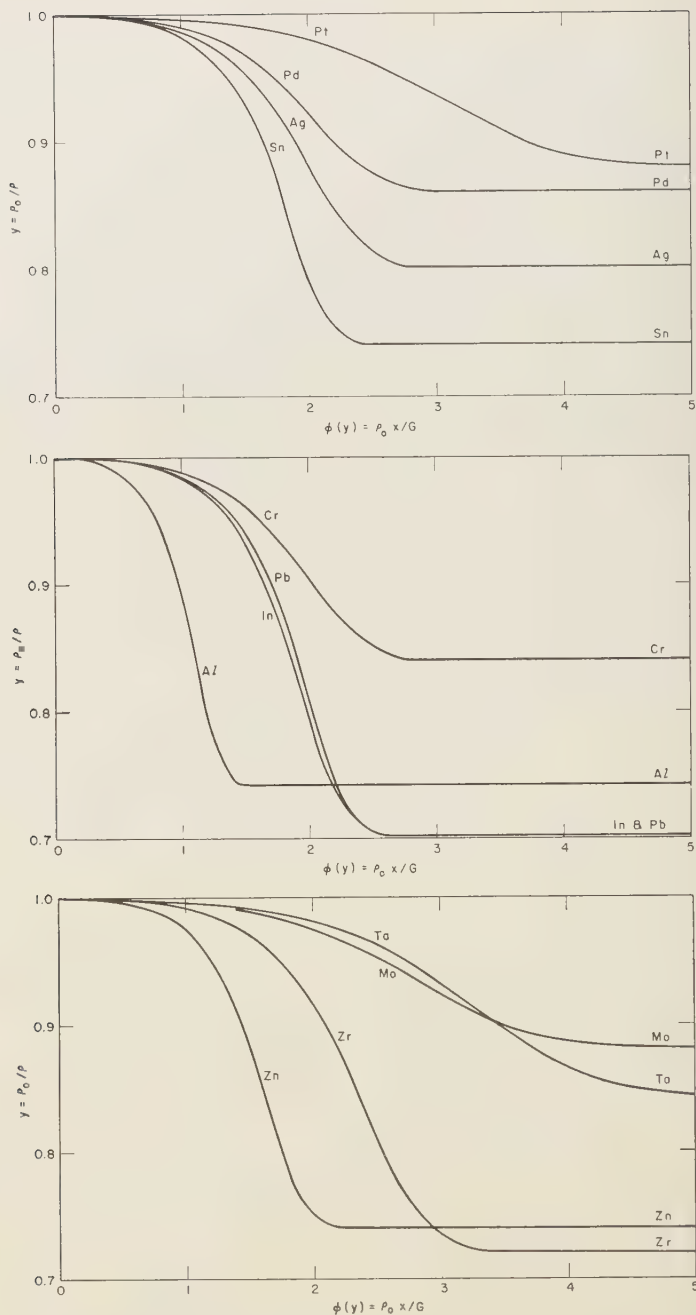
where D is the lattice spacing. Thus

$$G/\rho_0 = 10^5 S_m D \quad (32)$$

The values calculated in this way are given in the last line of Table 1 in stokes or poises divided by density in grams per cubic centimeter.

Feeble Shocks

A feeble shock will be defined as one so weak that its final compression falls at some point such as Q in Figure 1, and therefore is such that



Figs. 2, 3, and 4. Shock profiles for twelve metals giving the ratio of compression versus distance in units depending on the viscosity parameter (equation 25). These were each taken to maximum compression given in the data of *Walsh and others* [1957].

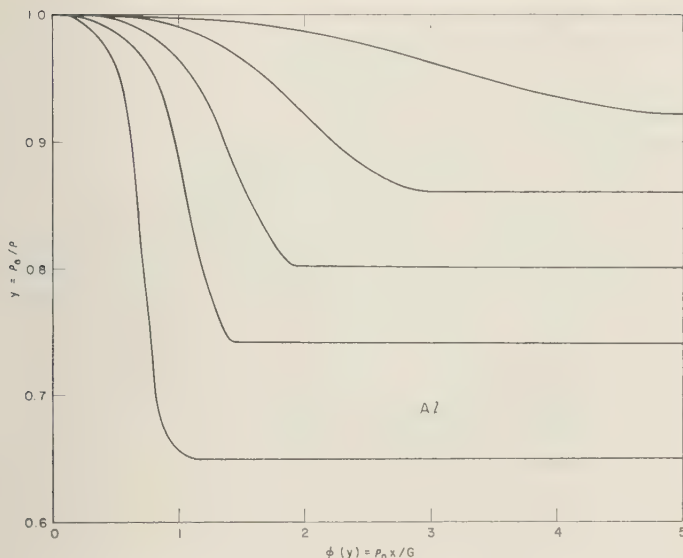


Fig. 5. Shock profiles for different final compressions of aluminum.

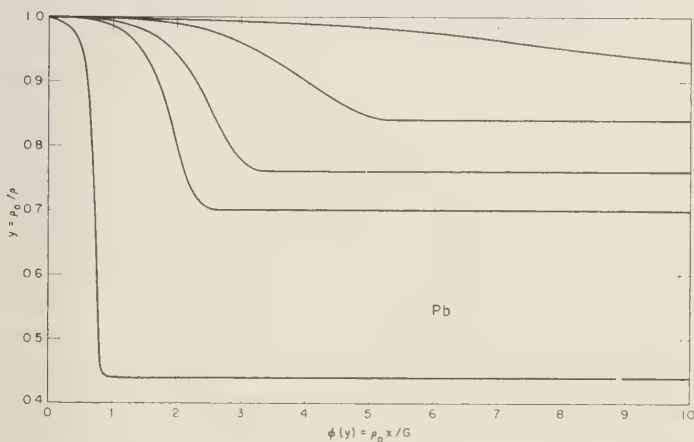


Fig. 6. Shock profiles for different final compressions of lead.

the negative stress p_0^* required by the Hugoniot jump conditions at the compression y_s is less than the negative stress p_s^* needed to produce yielding. This means that the permanent-regime solution we have found for the stronger shocks does not exist for the feeble shocks. Physically, what one would expect in this case is for the decrease in negative stress from zero to p_s^* to propagate as an elastic wave front with the speed of sound in the undisturbed material and

the remaining increase in negative stress from p_s^* to p_f^* at Q to propagate as a weak shock in the material in the wake of the elastic wave. If we therefore consider the particle velocity and compression behind the elastic wave and choose a reference system appropriately, we can look for a permanent-regime solution in the new reference system to represent the weak shock. The feeble shock is then split into an elastic front propagated ahead of a weak shock.

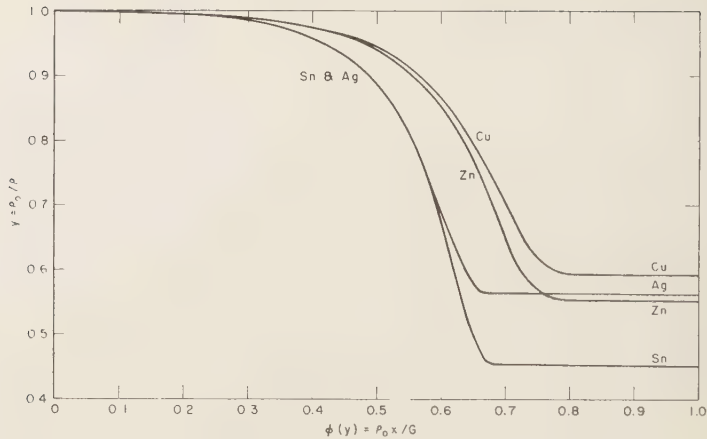


Fig. 7. Shock profiles for high compression of metals.

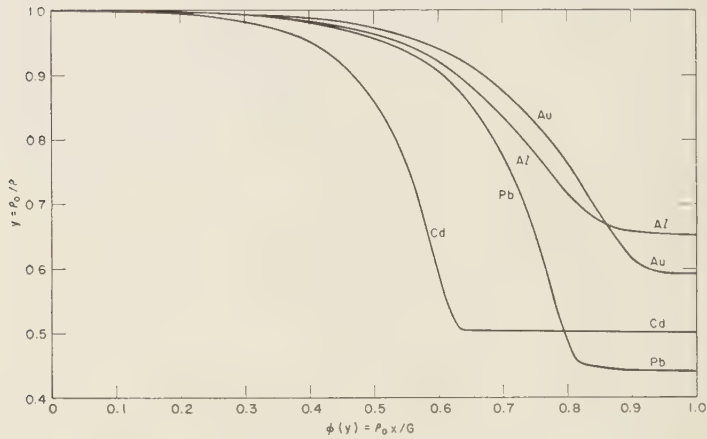


Fig. 8. Shock profiles for high compression of metals.

If we write C for the velocity of the elastic front, the particle velocity behind it is u_s , given by

$$\rho_0/\rho_s = (C - u_s)/C \tag{33}$$

We can now go back and repeat all the arguments concerning the permanent-regime solution, starting from ρ_s as the undisturbed density and taking the viscosity G as nonzero over the entire compression. The ratio $y = \rho_0/\rho$ must now be replaced by the ratio

$$\eta = \rho_s/\rho \tag{34}$$

and in place of (18) we would find

$$U_f'^2 = (p_f^* - p_s^*)/[\rho_s(1 - \rho_s/\rho_f)] \tag{35}$$

The actual shock speed would be $U_f' + u_s$ relative to the undisturbed material ahead of the elastic front. The negative stress p^* is defined by analogy with (1):

$$p^* = K(\eta) - \eta U_f' G (\partial \eta / \partial x) \tag{36}$$

TABLE 1. Values of G/ρ_0

	Al	Pb	Sn	Zn	Zr
S_0	0.014	0.009	0.009	0.012	0.012
S_m	25.	30.	30.	45.	28.
p	0.35	0.29	0.30	0.29	0.28
G/ρ_0	0.16	0.147	0.174	0.117	0.090

may assume that the elastic wave takes exactly when the negative stress reaches the value p_s^* , so that (36) yields

$$-U_f' G(\partial \eta / \partial x)_{\eta=1-0} = p_s^* - K(1) \quad (37)$$

the initial value of the compression gradient of the weak shock. The expression on the right (37) is just the stress difference due to the yielding process. The viscosity of the yielding process absorbs this difference in the shock profile. The analog of (25) for the profile is

$$1/G = \int_{\eta_s}^{\eta} \frac{(U_f - u_s) \eta' d\eta'}{(1 - \eta')[(U - u_s)^2 - (U_f - u_s)^2]} \quad (38)$$

where the velocities are all measured in the original reference system—they are relative to the undisturbed material ahead of the elastic front. The lower limit in this integral has been at the ratio

$$\eta_s = y_i/y_s$$

Figure 1, in order to remove the infinite 'tail' at the beginning of the profile. We can think of the yielding process as proceeding in the time interval between the passage of the elastic front and the arrival of the weak shock.

The modified equation giving the Hugoniot pressure in terms of the shock and particle velocities is easily derived:

$$p_f^* = p_s^* + \rho_s(U_f - u_s)(u_f - u_s) \quad (39)$$

where u_f is the final particle velocity relative to the original undisturbed material ahead of the elastic front, and u_s is given by (33):

$$u_s = C(1 - \rho_0/\rho_s)$$

Again U_f is the shock speed of the weak shock with respect to the undisturbed material ahead of the elastic wave front. Numerically, the difference between (39) and the usual equation where the splitting of the shock is neglected may generally be expected to be quite small; but for very weak shocks it can make an appreciable correction to the pressure calculation, and it has been included in recent experimental work on iron [Bancroft and others, 1956; Minshall, 1955] and aluminum [Doran and others, 1958]. We have not computed any profiles from (38).

DISCUSSION

The theory developed here gives shock profiles in terms of a single parameter that we have called the 'shear-yielding viscosity.' In fact, as was pointed out earlier in this paper, the use of a single viscosity parameter is a device which enables us to replace an actual solid by a highly simplified model whose equations of change can be solved exactly. In practice it must be expected that a considerable number of anelastic effects in the solid contribute to the determination of the structure of shock profiles, and we are not certain that all of them would behave in the way indicated in our equation (1). A more detailed discussion of various anelastic effects will be offered subsequently under the same general title.

For the present, however, it appears that the single viscosity parameter is not too unrealistic. By extrapolating the maximum slopes of the profiles as functions of the shock compressions to ideally complete compressions, and by relating the ultimate maximum slopes so derived to the lattice spacing of the material, we have obtained reasonable numerical estimates of the viscosity parameters.

Part II. Formal Theory of Anelasticity

INTRODUCTION

In Part I of this paper we presented a permanent-regime solution of the equations of motion of a shear-yielding solid and developed some theoretical profiles of shocks in a number of metals. In that work, we remarked that a more complete theory would require a detailed discussion of all anelastic effects in the material

in addition to the viscosity of the shear-yielding process. The theory of anelasticity developed by Zener [1948, chap. 7] applies immediately only to isotropic materials, and for our present purpose it is necessary to generalize that theory so as to apply it to crystalline materials. In the present paper we shall develop the theory for cubic crystals and derive the behavior of

small-amplitude sine waves in cubic crystals subject to anelastic effects. We shall not discuss shocks themselves, but the work is a necessary preliminary to such a discussion, to be given in subsequent papers.

Zener's standard one-dimensional anelastic solid is subject to the following linear relation between stress δ and strain e :

$$\delta + T \, d\delta/dt = M(e + T' \, de/dt) \quad (1)$$

where T and T' are relaxation times and M is the elastic modulus. This equation may be derived from two simple phenomenological relations:

$$\delta = Me - A\theta \quad (2a)$$

$$d\theta/dt = -\theta/T - B \, de/dt \quad (2b)$$

These relations can be understood as follows. In (2a), θ is the deviation from some standard value of a suitable potential of the anelastic mechanism (concentration of interstitial atoms, temperature, etc.), and A is the variation of pressure with respect to potential at constant volume:

$$A = (\partial p/\partial \theta)_V \quad (3a)$$

In (2b) the rate of relaxation of the anelastic potential, $-d\theta/dt$, is proportional to the potential deviation at constant volume, and B is the negative of the adiabatic variation of the potential with respect to volume:

$$B = -V(\partial \theta/\partial V)_s \quad (3b)$$

Note that the pressure is, by convention, negative stress, and relative volume increase is positive strain.

Elimination of θ between (2a) and (2b) yields (1), with

$$T' = T(1 + AB) \quad (4)$$

The relaxation time T may depend on the geometry of the situation. The relations (2) are purely formal, and for a more fundamental theory it is necessary to consider the anelastic mechanism in detail.

To obtain results applicable to shock propagation in solids, this paper gives a generalization of the formal theory to cubic crystals with

two independent relaxation mechanisms operating simultaneously.

THE BASIC EQUATIONS

In a cubic crystal, as distinct from an isotropic solid, a single parameter is insufficient to define the potential of any anelastic mechanism, particularly in a nonequilibrium situation. For example, the linear concentration of interstitial atoms may be different along different principal axes of the crystal; or the thermal energy associated with the normal modes may not be distributed isotropically among the three principal directions of propagation. Extremely high rates of change would be needed to cause a temperature anisotropy, but quite slow variations may be enough to cause anisotropy, for example, in concentration of dislocations. For this reason each anelastic mechanism will be characterized by three potentials, one associated with each principal axis in the crystal.

In order to simplify the algebra as much as possible, we shall assume that the principal axes of stress always coincide with the principal axes of the crystal structure. With this rather severe restriction it is possible to treat the three principal components of stress and the three principal components of strain as 3-component vectors. The elastic modulus is then a 3×3 matrix; the coefficients A and B of (2) also become 3×3 matrices, and the relaxation times can be written as 3×3 matrices. The generalizations of (2) become, in the appropriate notation,

$$\delta = M\mathbf{e} - A_1\theta_1 - A_2\theta_2 \quad (5a)$$

$$\theta_1 + T_1 \, \partial\theta_1/\partial t = -T_1 B_1 \, \partial\mathbf{e}/\partial t \quad (5b)$$

$$\theta_2 + T_2 \, \partial\theta_2/\partial t = -T_2 B_2 \, \partial\mathbf{e}/\partial t \quad (5c)$$

Roman type symbols M , etc., are 3×3 matrices, and boldface symbols are vectors; these latter are to be regarded as one-column matrices and multiplication is by the standard matrix rule. It is easy to see that these relations are invariant for coordinate rotations, so that the somewhat unusual interpretation of the components of stress, etc., as vector components is mathematically legitimate.

Due to crystal symmetry, each of the matrices M , T_n , A_n , and B_n have only two independent components; all have the form

$$X = \begin{bmatrix} X & x & x \\ x & X & x \\ x & x & X \end{bmatrix} \equiv [X; x] \quad (6)$$

all commute with each other on matrix multiplication. In general we shall use a lower-case letter for the off-diagonal component and a capital letter for the diagonal component of a matrix.

The algebraic problem is to eliminate all the components of both potentials θ_1 and θ_2 from (1) and derive the analog of Zener's relation, (1).

THE ANELASTIC RELATIONS

We define the matrix operators

$$n = 1, 2 \quad D_n = E + T_n \partial/\partial t \quad (7)$$

where E is the identity 3×3 matrix. These operators commute with each other and with any constant matrix of the form (6). Thus by operating on (5a) by $D_1 D_2$ and making use of (5b) and (5c) we obtain at once the desired relation, the generalization of (1):

$$D_2 \delta = D_1 D_2 M e \\ + (D_2 A_1 B_1 T_1 + D_1 A_2 B_2 T_2) \partial e / \partial t \quad (8)$$

In its raw form, this equation is not very useful; we wish rather to express it explicitly in terms of its three components. To do this it is necessary to go through some tedious algebra, which can be greatly facilitated by using the notation suggested in (6). Thus we have the following simple theorems:

$$Y = [XY + 2xy; Xy + xy + Yx] \quad (9)$$

$$X + Y = [X + Y; x + y] \quad (10)$$

$$X \delta = \begin{bmatrix} X \delta_1 + x(\delta_2 + \delta_3) \\ X \delta_2 + x(\delta_3 + \delta_1) \\ X \delta_3 + x(\delta_1 + \delta_2) \end{bmatrix} \quad (11)$$

the last being a column vector. To save space, we shall in the sequel write column vectors as rows of components:

$$\delta = (\delta_1, \delta_2, \delta_3)$$

The first two terms of (8) are comparatively simple; they become

$$\begin{aligned} D_1 D_2 \delta &= (\delta_1, \delta_2, \delta_3) \\ &+ (T_1 + T_2)(\dot{\delta}_1, \dot{\delta}_2, \dot{o}_3) + (t_1 + t_2) \\ &\quad \cdot (\dot{\delta}_2 + \dot{\delta}_3, \dot{\delta}_3 + \dot{\delta}_1, \dot{\delta}_1 + \dot{\delta}_2) \\ &+ (T_1 T_2 + 2t_1 t_2)(\ddot{\delta}_1, \ddot{\delta}_2, \ddot{\delta}_3) \\ &+ (T_1 t_2 + T_2 t_1 + t_1 t_2) \\ &\quad \cdot (\ddot{\delta}_2 + \ddot{\delta}_3, \ddot{\delta}_3 + \ddot{\delta}_1, \ddot{\delta}_1 + \ddot{\delta}_2) \end{aligned} \quad (12)$$

$$\begin{aligned} D_1 D_2 M e &= M(e_1, e_2, e_3) \\ &+ m(e_2 + e_3, e_3 + e_1, e_1 + e_2) \\ &+ \{M(T_1 + T_2) + 2m(t_1 + t_2)\} \\ &\quad \cdot (\dot{e}_1, \dot{e}_2, \dot{e}_3) + \{M(t_1 + t_2) \\ &\quad + m(T_1 + t_1 + T_2 + t_2)\} \\ &\quad \cdot (\dot{e}_2 + \dot{e}_3, \dot{e}_3 + \dot{e}_1, \dot{e}_1 + \dot{e}_2) \\ &+ \{M(T_1 T_2 + 2t_1 t_2) \\ &\quad + 2m(T_1 t_2 + T_2 t_1 + t_1 t_2)\} \\ &\quad \cdot (\ddot{e}_1, \ddot{e}_2, \ddot{e}_3) + \{(T_1 T_2 + 2t_1 t_2)m \\ &\quad + (T_1 t_2 + T_2 t_1 + t_1 t_2)(M + m)\} \\ &\quad \cdot (\ddot{e}_2 + \ddot{e}_3, \ddot{e}_3 + \ddot{e}_1, \ddot{e}_1 + \ddot{e}_2) \end{aligned} \quad (13)$$

The dots, as usual, represent time derivatives.

For the last two terms of (8) it is convenient to introduce the following abbreviations:

$$\left. \begin{aligned} n &= 1, 2: \\ K_n &= A_n B_n + 2a_n b_n \\ J_n &= A_n b_n + B_n a_n + a_n b_n \\ K_n - J_n &= (A_n - a_n)(B_n - b_n) \end{aligned} \right\} \quad (14)$$

We then find the following relation:

$$\begin{aligned} (D_2 A_1 B_1 T_1 + D_1 A_2 B_2 T_2) \dot{e} &= \Sigma_n (K_n T_n + 2J_n t_n)(\dot{e}_1, \dot{e}_2, \dot{e}_3) \\ &+ \Sigma_n \{K_n t_n + (T_n + t_n)J_n\} \\ &\quad \cdot (\dot{e}_2 + \dot{e}_3, \dot{e}_3 + \dot{e}_1, \dot{e}_1 + \dot{e}_2) \\ &+ \{\Sigma_n K_n (T_1 T_2 + 2t_1 t_2) \\ &\quad + 2\Sigma_n J_n (T_1 t_2 + T_2 t_1 + t_1 t_2)\} \\ &\quad \cdot (\ddot{e}_1, \ddot{e}_2, \ddot{e}_3) + \{\Sigma_n K_n (T_1 t_2 + T_2 t_1 + t_1 t_2) \end{aligned}$$

$$+ \Sigma_n J_n (T_1 T_2 + 3 t_1 t_2 + T_1 t_2 + T_2 t_1) \} \quad T_n + 2 t_n = T_{pn}, \quad T_n - t_n = T_{sn} \quad \left. \begin{array}{l} K = (M + 2m)/3, \quad 2\mu = M - m \\ K_n' = (M + 2m + K_n + 2J_n)/3 \\ K'' = (M + 2m + \Sigma_n K_n + 2\Sigma_n J_n)/3 \\ 2\mu_n' = M - m + K_n - J_n \\ 2\mu'' = M - m + \Sigma_n K_n - \Sigma_n J_n \end{array} \right\} \quad (20)$$

$$\cdot (\ddot{e}_2 + \ddot{e}_3, \ddot{e}_3 + \ddot{e}_1, \ddot{e}_1 + \ddot{e}_2) \quad (15)$$

Collecting (12), (13), and (15), we have the component form of (8).

It is convenient now to define the mean stress δ and the volume dilation e as

$$\left. \begin{array}{l} \delta = (\delta_1 + \delta_2 + \delta_3)/3 \\ e = e_1 + e_2 + e_3 \end{array} \right\} \quad (16)$$

The α th component of (8) then can be written out in detail: call the result equation (C). We can then add together all three such equations; the result is a relation giving the mean stress as a function of the volume dilation. If we introduce the two pairs of scalar differential operators defined by

$$\left. \begin{array}{l} n = 1, 2 \quad D_{pn} = 1 + (T_n + 2t_n)\partial/\partial t \\ n = 1, 2 \quad D_{sn} = 1 + (T_n - t_n)\partial/\partial t \end{array} \right\} \quad (17)$$

then the mean stress/volume relation reduces to

$$\begin{aligned} D_{p1} D_{p2} \delta &= (1/3)(M + 2m)e + (1/3)\Sigma_n \\ &\cdot \{ (M + 2m + K_n + 2J_n)(T_n + 2t_n) \} \dot{e} \\ &+ (1/3) \{ M + 2m + \Sigma_n (K_n + 2J_n) \} \\ &\cdot (T_1 + 2t_1)(T_2 + 2t_2) \ddot{e} \end{aligned} \quad (18)$$

If we subtract this equation from equation (C), mentioned above, the result is a relation giving the shear stress relation: $\alpha = 1, 2, 3$

$$\begin{aligned} D_{s1} D_{s2} (\delta_\alpha - \delta) &= (M - m)(e_\alpha - e/3) \\ &+ \Sigma_n (M - m + K_n - J_n)(T_n - t_n) \\ &\cdot (\dot{e}_\alpha - \dot{e}/3) \\ &+ (M - m + \Sigma_n K_n - \Sigma_n J_n) \\ &\cdot (T_1 - t_1)(T_2 - t_2)(\ddot{e}_\alpha - \ddot{e}/3) \end{aligned} \quad (19)$$

We can now eliminate the mean stress from the last two equations and so derive a differential equation containing only a single component of stress. It is convenient to abbreviate the coefficients further:

Let

$$\begin{aligned} D_{p1} D_{p2} D_{s1} D_{s2} \delta_\alpha &= K D_{s1} D_{s2} e \\ &+ 2\mu D_{p1} D_{p2} (e_\alpha - e/3) \\ &+ \Sigma_n K_n' T_{pn} D_{s1} D_{s2} \dot{e} \\ &+ 2\Sigma_n \mu_n' T_{sn} D_{p1} D_{p2} (\dot{e}_\alpha - \dot{e}/3) \\ &+ K'' T_{p1} T_{p2} D_{s1} D_{s2} \ddot{e} \\ &+ 2\mu'' T_{s1} T_{s2} D_{p1} D_{p2} (\ddot{e}_\alpha - \ddot{e}/3) \end{aligned} \quad (21)$$

This is the form of anelastic relation that we need in order to be able to write the dynamic equation of motion of an element of material and we proceed at once in the next section to do this.

THE ANELASTIC EQUATIONS OF MOTION

Let (u_1, u_2, u_3) be the relative displacement vector, so that

$$e_\alpha = \partial u_\alpha / \partial x_\alpha, \quad \alpha = 1, 2, 3 \quad (22)$$

and

$$\partial \delta_\alpha / \partial x_\alpha = \rho \ddot{u}_\alpha \quad (23)$$

is the equation of motion [Band, 1959]. Consider irrotational waves in which a linear compression occurs parallel to the first axis: $e = e$, $e_2 = e_3 = 0$, and e_1 is a function only of x and t . Then from (21) we derive at once the wave equation for plane compression waves:

$$\begin{aligned} \rho D_{p1} D_{p2} D_{s1} D_{s2} \ddot{u} &= \{ K D_{s1} D_{s2} \\ &+ (4/3)\mu D_{p1} D_{p2} \} \partial^2 u / \partial z^2 \\ &+ \{ \Sigma_n K_n' T_{pn} D_{s1} D_{s2} \\ &+ (4/3)\Sigma_n \mu_n' T_{sn} D_{p1} D_{p2} \} \partial^2 \dot{u} / \partial z^2 \\ &+ \{ K'' T_{p1} T_{p2} D_{s1} D_{s2} \\ &+ (4/3)\mu'' T_{s1} T_{s2} D_{p1} D_{p2} \} \partial^2 \ddot{u} / \partial z^2 \end{aligned} \quad (24)$$

the solenoidal wave is not so directly derived because the stress and strain have to be in normal form, and therefore the direction of propagation is at 45° to two principal axes, and the relative displacement is also at 45° to the axes. If u is the magnitude of the relative displacement vector and z is the direction of propagation, the principal components of stress are

$$\frac{1}{2} \partial u / \partial z, \quad 0, \quad -\frac{1}{2} \partial u / \partial z \quad (25)$$

We also have the transformations

$$(1/2^{1/2})(\partial/\partial x_1 - \partial/\partial x_2) = \partial/\partial z$$

and

$$(1/2^{1/2})(u_1 + u_2) = u \quad (26)$$

Combining two of equations (21) we can then derive the following wave equation for transverse waves:

$$\begin{aligned} D_{p1} D_{p2} D_{s1} D_{s2} \ddot{u} &= \mu D_{p1} D_{p2} \partial^2 u / \partial z^2 \\ &+ \Sigma_n \mu_n' T_{sn} D_{p1} D_{p2} \partial^2 \dot{u} / \partial z^2 \\ &+ \mu'' T_{s1} T_{s2} D_{p1} D_{p2} \partial^2 \ddot{u} / \partial z^2 \end{aligned} \quad (27)$$

Equations (24) and (27) reduce to the well known elastic wave equations when the relaxation times all vanish and the dots are interpreted as partial time derivatives.

Solutions of the anelastic wave equations can be found of the form

$$u = u_0 \exp(i\omega t + ikz) \quad (28)$$

which, if we make ω real and allow k to be complex, represents a forced oscillation with space damping during propagation. If we make ω real and allow w to be complex, the solution presents standing waves with time damping. An arbitrary wave form can be built up by Fourier integrals. Because the wave equations are linear, we can also use the method of the Laplace transform to describe the propagation of any pulse form and study its attenuation. Since these methods are standard [Morse and Feshbach, 1953, p. 1339], there is no point in repeating them in detail in the present paper. We shall restrict ourselves to some qualitative comments on the physical interpretation of the equations.

The relaxation times T_{p1} and T_{p2} are associated with processes stimulated by hydrostatic compression, as is apparent from (18), and the times T_{s1} and T_{s2} are associated with shear strains (see equation 19). From the structure of (24) and (27) it is easy to see that there are in general four frequencies where attenuation would be a maximum for either type of wave, but there are significant differences between the solenoidal and the irrotational waves, respectively. If the frequency ω is set so high that all factors ωT are much greater than unity, the wave equations reduce, respectively, to the following:

$$\left. \begin{aligned} &\text{longitudinal:} \\ &\rho \ddot{u} = (1/\omega^2)(K'' + 4\mu''/3) \partial^2 \ddot{u} / \partial z^2 \\ &\text{transverse:} \\ &\rho \ddot{u} = (1/\omega^2)\mu'' \partial^2 \ddot{u} / \partial z^2 \end{aligned} \right\} \quad (29)$$

For sinusoidal waves, therefore, the modified elastic constants K'' and μ'' take over at very high frequencies—the so-called completely unrelaxed coefficients.

At intermediate frequencies we consider two illustrative cases. Suppose that T_{p1} is the shortest of the relaxation times, and that a high frequency is chosen so that $\omega T_{p1} \sim 1$, and all other $\omega T \gg 1$. One can show that the longitudinal wave is strongly damped at such frequencies, and that the transverse waves suffer negligible damping. If in the same material T_{s2} happens to be the longest of the relaxation times, and a low frequency is chosen so that $\omega T_{s2} \sim 1$, all other $\ll 1$, it turns out that both types of wave suffer strong damping. For widely separated relaxation times, the transverse wave shows only three attenuation peaks and the longitudinal wave shows four.

On the other hand, if T_{s1} happens to be the shortest of the relaxation times and the others are all widely separated, both types of wave are attenuated strongly at frequencies near $1/T_{s1}$, and the transverse wave exhibits all four attenuation peaks. In general, longitudinal waves show strong damping peaks near the frequencies $1/T_{pn}$, and weak peaks near $1/T_{sn}$; the transverse wave case is reversed. A comparison of the attenuation spectra of transverse and longitudinal waves should prove a useful approach to a study of relaxation mechanisms.

THERMOELASTIC RELATIONS

In general, it is not realistic to neglect the effects of thermoelastic coupling. In some cases it is possible simply to interpret the formalism developed above in such a way that this coupling is taken care of. Thus the potential θ_1 could be interpreted as temperature, and, for slow enough changes, all three components would be tied together, i.e., the off-diagonal elements of the temperature relaxation time matrix would be effectively zero. However, with this interpretation it is no longer true that the simple relaxation equation (5b) is generally adequate to describe the known physical behavior of thermal diffusion. The generally valid relation superseding (5b) can be derived from the conservation of thermal energy. Let the thermal conductivity be χ , the heat capacity per gram C_v , the temperature deviation θ , and the absolute temperature θ . Then the conservation of energy E per unit volume is

$$\chi \nabla^2 \theta = \rho C_v \partial \theta / \partial t + (\partial E / \partial V)_s (\partial V / \partial t) + p \partial V / \partial t$$

Assuming internal equilibrium of thermal energy, we see that the thermodynamic identity holds:

$$p + (\partial E / \partial V)_s = \theta (\partial p / \partial \theta)_v$$

and the conservation equation takes the form

$$\left. \begin{aligned} g^2 \nabla^2 \theta - \partial \theta / \partial t &= q \partial e / \partial t \\ \text{where} \\ g^2 &= \chi / \rho C_v, \quad q = \theta (\partial p / \partial \theta)_v / \rho C_v \\ &= \theta \gamma \end{aligned} \right\} \quad (30)$$

where γ is Grüneisen's ratio. This reduces to the simple relaxation equation only if

$$g^2 \nabla^2 \theta = -\theta / T$$

Zener [1948, p. 76] used this relation to develop his theory of orthogonal potentials, which is applicable when well-defined boundary conditions are given. It is, however, not convenient to use this method in general, and we proceed to set up and solve the appropriate general equations for the cubic crystals with undefined boundary conditions.

We introduce the operators

$$D_t = g^2 \nabla^2 - \partial / \partial t, \quad D = E + T \partial / \partial t \quad (31)$$

where T is the relaxation-time matrix for the second mechanism (nonthermal), and we let the potential for the second mechanism be represented by the vector ϕ . The vectors $\theta = (\theta, \theta, \theta)$, $q = (q, q, q)$ are also useful. Again we write $A = (\partial p / \partial \theta)_s$ in place of A_1 , and B in place of A_2 ; then the three equations analogous to (5) are

$$\delta = M e - A \theta - B \phi \quad (32a)$$

$$D_t \theta = q \partial e / \partial t \quad (32b)$$

$$D \phi = -\Gamma \partial e / \partial t \quad (32c)$$

Here the matrix Γ replaces $T_2 B_2$ of (5c).

If we operate on (32a) with $D_t D$, on (32b) with D , and on (32c) with D_t , a simple linear combination of the results eliminates the potentials ϕ and θ to yield the anelastic relation between stress and strain, analogous to (8):

$$D_t D \delta = D_t D M e - D q A \partial e / \partial t + D_t B \Gamma \partial e / \partial t \quad (33)$$

Again we note that all constant matrices and operators commute with each other.

The α th component of this equation is

$$\begin{aligned} D_t \{ (1 + T \partial / \partial t) \delta_\alpha + t (\partial / \partial t) (3 \delta - \delta_\alpha) \} \\ = D_t \{ [M + (MT + 2mt) \partial / \partial t] e_\alpha \\ + [m + (Mt + mT + mt) \partial / \partial t] (e - e_\alpha) \\ - A [1 + (T + 2t) \partial / \partial t] q \partial e / \partial t \\ + D_t \{ (B\Gamma + 2b\gamma) \partial e_\alpha / \partial t \\ + (B\gamma + b\Gamma + 2b\gamma) (\partial / \partial t) (e - e_\alpha) \} \end{aligned} \quad (34)$$

If we add the three components, $\alpha = 1, 2, 3$ and rearrange terms, we find

$$D_t D_p \delta = D_t K D_p e - A q D_p \dot{e} + D_t J_p \dot{e} \quad (35)$$

where

$$\left. \begin{aligned} D_p &= 1 + (T + 2t) \partial / \partial t \\ K &= (M + 2m) / 3 \\ J_p &= (B + 2b)(\Gamma + 2\gamma) / 3 \end{aligned} \right\} \quad (36)$$

This is the anelastic equation for hydrostatic compressions.

Subtracting (35) from (34) gives the anelastic equation for shears:

$$D_s(\delta_\alpha - \delta) = 2\mu D_t D_s(e_\alpha - e/3) \\ + D_t J_s(\dot{e}_\alpha - \dot{e}/3) \quad (37)$$

re

$$\left. \begin{aligned} D_s &= 1 + (T - t) \partial/\partial t \\ 2\mu &= M - m \\ J_s &= (B - b)(\Gamma - \gamma) \end{aligned} \right\} \quad (38)$$

note that the operator D_t could be canceled in (37) leaving an arbitrary constant. Operation (35) with D_s and on (37) with D_p and using the results, we find the anelastic equation a single component of the stress:

$$D_p D_s \delta_\alpha = K D_t D_p D_s e \\ + 2\mu D_t D_p D_s(e_\alpha - e/3) \\ - Aq D_p D_s \dot{e} + D_t D_s J_p \dot{e} \\ + D_t D_p J_s(\dot{e}_\alpha - \dot{e}/3) \quad (39)$$

in this equation we can derive the linear relations of wave propagation as before and consider the effects of thermoelastic coupling in both compression and shear waves.

THE THERMOELASTIC EQUATIONS OF MOTION

Using the same equations of motion as before (equations 22 and 23), we can convert (39) to the propagation equation for compression waves:

$$D_t D_p D_s \ddot{u} \\ = (K + 4\mu/3) D_t D_p D_s \partial^2 u / \partial z^2 \\ - Aq D_p D_s \partial^2 \dot{u} / \partial z^2 \\ + D_t D_s J_p \partial^2 \dot{u} / \partial z^2 \\ + (2/3) D_t D_p J_s \partial^2 \dot{u} / \partial z^2 \quad (40)$$

and, by using the argument around (25) and (26), again, we can derive the propagation equation for transverse waves:

$$D_t D_p D_s \ddot{u} = \mu D_t D_p D_s \partial^2 u / \partial z^2 \\ + \frac{1}{2} D_t D_s J_p \partial^2 \dot{u} / \partial z^2 \quad (41)$$

note again that the operator D_t may be canceled from the last equation, leaving an arbitrary constant; thermal diffusion has no effect on the attenuation of transverse waves.

The interesting effects of thermal flow are confined to the behavior of compression waves.

Interpreting the dots as partial time derivatives, we see that solutions of the wave equation can again be found in the form

$$u = u_0 \exp(i\omega t + ikz)$$

when the operators become algebraic multipliers,

$$\left. \begin{aligned} D_p &= 1 + i\omega T_p, \quad D_s = 1 + i\omega T_s \\ D_t &= -k^2 g^2 - i\omega \end{aligned} \right\} \quad (42)$$

The substitution of these expressions into (40) yields the following equation between k and ω :

$$\rho \omega^2 / k^2 = K + 4\mu/3 \\ + i\omega \{ Aq / (k^2 g^2 + i\omega) \\ + J_p / (1 + i\omega T_p) \\ + 2J_s / 3(1 + i\omega T_s) \} \quad (43)$$

The ratio ω^2/k^2 is the square of the complex velocity of propagation, C^2 , and we can, in the smaller terms, write $k^2 = \omega^2/C_0^2$, where $C_0^2 = (K + 4\mu/3)/\rho$. To a good approximation, then, (43) becomes

$$\rho C^2 = \rho C_0^2 + Aq / (1 - i\omega T_i) \\ + i\omega \{ J_p / (1 + i\omega T_p) \\ + 2J_s / 3(1 + i\omega T_s) \} \quad (44)$$

where

$$T_i = g^2 / C_0^2 \quad (45)$$

is the relaxation time for thermal diffusion.

We suppose now that the times T_p and T_s are long compared with T_i ; then at least three ranges of frequency are of interest, namely, very high ones for which the ωT 's are all much larger than unity, intermediate ones for which ωT_i is less than unity, but ωT_p and ωT_s are still larger than unity, and low ones for which all the ωT 's are less than unity. The equation for the complex velocity of propagation (44) reduces in these cases respectively to

$$\text{high } \omega: \quad \rho C^2 = \rho C_0^2 + J_p / T_p \\ + 2J_s / 3T_s + iAq / \omega T_i \quad (46)$$

$$\text{int. } \omega: \quad \rho C^2 = \rho C_0^2 + Aq + J_p / T_p \\ + 2J_s / 3T_s + i\omega Aq T_i \quad (47)$$

$$\text{low } w: \quad \rho C^2 = \rho C_0^2 + Aq \\ + iw(J_p + 2J_s/3 + AqT_i). \quad (48)$$

It is evident that the effect of the thermal term on the velocity of propagation is the opposite of the effects of nonthermal relaxation mechanisms; increasing the frequency through $1/T_p$ or $1/T_s$ increases the velocity of propagation, whereas increasing it through $1/T_i$ decreases the velocity of propagation to the so-called isothermal value [Mason, 1950 p. 479]. Approximate values of the dissipation function can be read from the imaginary parts of the expressions in (46) to (48):

$$\left. \begin{aligned} \text{high } w: \\ 1/Q &= Aq/\rho C_0^2 wT_i \\ \text{int. } w: \\ 1/Q &= wAqT_i'/\rho C_0 \\ \text{low } w: \\ 1/Q &= w(J_p + 2J_s/3 + AqT_i)/\rho C_0 \end{aligned} \right\} \quad (49)$$

The fact that the effects of thermal flow are so distinct from those of the nonthermal relaxation mechanism makes it quite essential to deal with the thermal problem explicitly, as we have done here, rather than treating it formally as a relaxation process in the way mentioned at the beginning of the last section.

DISCUSSION

The results obtained here have been explicitly limited to cubic crystals and to plane disturbances propagated along a principal axis. It is obvious that, although not applicable to materials of less than cubic symmetry, the results are valid for propagation of plane waves in isotropic materials. The question arises whether they could be applied to homogeneous polycrystalline materials in which microcrystals are randomly oriented. It would seem that such a material could be treated as macroscopically isotropic and that the microstructure provides exactly the kind of mechanism that can be described in terms of the relaxation term represented in (32) by the potential ϕ and discussed in some detail by Zener [1948, p. 84].

Indeed, this opens up the possibility of discussing the combined effects of a macroscopic

thermal diffusion and a grain-structure thermal relaxation on the propagation of small-amplitude vibrations. We hope to return to this later.

Generalization of the theory to single crystals with less than cubic symmetry, or to disturbances propagated in cubic crystals in other than the principal directions, would involve formidable complications in notation and will not be attempted.

In order to apply the results of the present investigation to the study of shock waves in solids it is essential first to abandon the approximation to linearity. All the partial time derivatives, appearing particularly in equations (4), (7), (17), (30), (31), (32), (33), and (34) must be rewritten as convective derivatives $d/dt = \partial/\partial t + \mathbf{v} \cdot \text{grad}$. The equations then become completely intractable, in part because the time derivatives no longer commute with the space derivatives. Fortunately, in the permanent-regime problem, where the partial time derivatives vanish, some degree of commutability is restored and the equations can sometimes be solved. We report work along these lines in Part III.

There are other, more physical, difficulties involved in discussing applications to shock propagation. The shock front may rise in time comparable to, or even much shorter than T_i , the relaxation time for thermal diffusion. Therefore, the assumption of internal thermal equilibrium made in deriving (30) and implied in (32b) is insecure. To make this point more clear: when we analyze thermal energy into the energy of elastic vibrations—normal modes or phonons—the dominant contribution, in fact, comes from modes having frequencies in the neighborhood of $1/T_i$. Superposing low-frequency sound vibrations on such a distribution produces a second-order effect on the high-frequency components, modulating their frequencies and so causing variations in temperature.

But superposing a very high frequency vibration, comparable in frequency to $1/T_i$, can have no such effect; it simply adds its own contribution to the thermal energy and so disturbs the internal equilibrium of the energy distribution. Any acoustical disturbance, such as a shock wave containing an appreciable proportion of its energy in sufficiently high frequency modes can be regarded as disturbing the internal

ilibrium of the thermal energy at any point, addition to the modulating effect of its low-viscosity components.

The anelastic effects of temperature modulations are adequately taken care of by the diffusion

theory incorporated in the last two sections; the anelastic effects of disturbances in the distribution of energy have not been hitherto considered, and we shall take them up in a later paper.

Part III. Steady-State Compressional Profiles Propagating in Heat-Conducting Anelastic Solids

INTRODUCTION

In Part I of this paper we discussed the permanent-regime profiles of compressive shocks in a shear-yielding solid by means of a single viscosity-type parameter. In Part II we developed a general theory of the anelasticity of solids with cubic crystal symmetry and applied it to the theory of small-amplitude waves, that is, the linear approximation to the equations of motion. In this part the results of Part II will be applied to the nonlinear equations of motion and the profiles of permanent-regime or steady-state solutions investigated under various conditions.

As in Part I, the shear-yielding process will be represented by the single viscosity-type parameter, and the elastic region below the yield point will be neglected. From the discussion in Part I it is clear that for sufficiently strong shocks such neglect is permissible; the small strain in which the stress is less than the yield stress makes almost no difference to the profile. For weak shocks, of course, this neglect would be serious, and we must explicitly disclaim the applicability of the results of the present paper to weak shocks.

In Part I we ignored the effects of temperature which are, of course, essentially anelastic. This meant, in effect, that the permanence of the relations we found was somewhat provisional. It is probable that such profiles would lead to gradual accumulation of heat or, vice versa, to a gradual reduction in temperature of the compressed material; we could hardly proceed without feeding the equations governing thermal flow into the equation of the profile. Here we shall be able to attack the problem by taking the temperature taken into account from the beginning and seek a permanent-regime solution in which temperature is also a steady

variable. The existence of such solutions has been proved very generally [Gilburg, 1951; Drummond, 1958]. We shall approach such a solution by the method of successive approximations.

THE GENERAL ANELASTIC EQUATIONS

Shear-yielding viscosity was omitted from the general anelastic relations developed in Part II; here we shall revise those equations to include this viscosity. Also, for future convenience when we come to discuss the relaxation effects of phonon scattering and of dislocations, it is as well to separate the stress into two parts: (a) the configurational stress matrix δ' and (b) the dynamical stress matrix δ'' . The configurational stress is defined as the stress due entirely to the intermolecular potential energy; it is the stress that would exist if the material were to be cooled to 0°K without altering its configuration or strain. The dynamical stress is the difference between the total stress and the configurational stress, sometimes called the 'phonon pressure,' and is due to the kinetic energy of the molecules. This separation is familiar in the Grüneisen treatment of the solid state [Born and Huang, 1954, p. 49].

The viscosity and the anelastic relaxations affect only the configurational stress, and we write

$$\delta' = M'e - B\phi + \eta\dot{\epsilon} \quad (1)$$

where M' is the configurational part of the elastic modulus matrix and η is the shear-yielding viscosity matrix. The dynamical stress depends on strain and temperature, and we write

$$\delta'' = M''e - A\theta \quad (2)$$

where M'' is the dynamical part of the elastic

modulus matrix. Otherwise, the notation is the same as that in Part II, equation (32).

By adding these two equations we arrive at an equation identical with (32a) of Part II, except for the extra viscosity term.

$$\delta = M\epsilon - A\theta - B\phi + \eta d\epsilon/dt \quad (3a)$$

$$D_t\theta = q d\epsilon/dt \quad (3b)$$

$$D\phi = -\Gamma d\epsilon/dt \quad (3c)$$

where for the last two equations we have simply repeated the corresponding equations (32b) and (32c) of Part II. To avoid complications, the source of heat due to the work done against the shear-yielding viscosity is neglected in (3b); a perturbation method of estimating this and making the appropriate corrections is given in the last section of this paper.

For the permanent-regime solutions it is not sufficient to take the partial time derivatives; in each of the above equations the correct derivative is the convective one:

$$d/dt = \partial/\partial t + \mathbf{v} \cdot \nabla \quad (4)$$

where \mathbf{v} is the velocity of the material. The operators are then

$$\begin{aligned} D &= E + T d/dt \\ &= E + T(\partial/\partial t + \mathbf{v} \cdot \nabla) \end{aligned} \quad (5)$$

$$D_t = g^2 \nabla^2 - d/dt$$

$$= g^2 \nabla^2 - \partial/\partial t - \mathbf{v} \cdot \nabla$$

The operator D_t no longer commutes with D because the velocity \mathbf{v} is not a constant vector. The operator d/dt and D do commute with each other, however. In order to eliminate θ and ϕ from equations (3) it is now necessary to postulate the existence of an inverse operator D_t^{-1} such that (3b) becomes

$$\theta = q D_t^{-1} d\epsilon/dt \quad (6)$$

Simple operations on (3) then yield the following anelastic relations:

$$\begin{aligned} D\delta &= DME - A Dq D_t^{-1} d\epsilon/dt \\ &\quad + B\Gamma d\epsilon/dt + \eta D d\epsilon/dt \end{aligned} \quad (7)$$

This equation replaces (33) of Part II and can similarly be separated into its individual com-

ponents:

$$\alpha = 1, 2, 3;$$

$$\begin{aligned} D_p D_s \delta_\alpha &= D_p D_s K e \\ &\quad + 2\mu D_p D_s (e_\alpha - e/3) \\ &\quad + D_p D_s \eta_p d\epsilon/dt \\ &\quad + 2\eta_s D_p D_s d(e_\alpha - e/3)/dt \\ &\quad - Aq D_p D_s D_t^{-1} d\epsilon/dt \\ &\quad + D_s J_p d\epsilon/dt + D_p J_s d(e_\alpha - e/3)/dt \end{aligned}$$

where η_p and η_s are the volume and shear viscosities, respectively.

In Part I, equation (4), we set down the equations of motion without making the linear approximation

$$\partial(\rho v)/\partial t + \partial(\rho v^2)/\partial x = \partial \delta_1/\partial x$$

For the permanent-regime equation we $\partial/\partial t = 0$ both here and in (4); also we have from (7) of Part I,

$$v = Uy, \quad y = \rho_0/\rho \quad (8)$$

where U is the velocity of propagation of the profile. We also have $e_1 = y - 1$ and $e = e_1$ the linear compression. The propagation equation derived from (8) is then easily obtained thus from (9)

$$\partial(y\rho_0 U^2)/\partial x = \partial \delta_1/\partial x$$

which integrates at once to

$$(y - 1)\rho_0 U^2 = \delta_1 \quad (9)$$

We define

$$D \equiv d/dt = yU \partial/\partial x \quad (10)$$

Equation (8) becomes

$$\begin{aligned} \rho_0 U^2 (1 + T_p D)(1 + T_s D)(y - 1) \\ = (1 + T_p D)(1 + T_s D) \\ \cdot [(K + 4\mu/3)(y - 1)] \\ + (1 + T_p D)(1 + T_s D) D \\ \cdot [(\eta_p + 4\eta_s/3)(y - 1)] \\ + [J_p(1 + T_s D) + (2J_s/3) \\ \cdot (1 + T_p D)] D(y - 1) \end{aligned}$$

$$- Aq(1 + T_p D) \\ \cdot (1 + T_s D) D_t^{-1} D(y - 1) \quad (13)$$

operator D_t^{-1} here causes trouble because it does not commute with the other operators; it will have to be treated approximately and each case examined separately.

ADIABATIC APPROXIMATION, HOOKEAN SOLID

In the completely relaxed case. We consider the effects of yielding viscosity and thermal conductivity alone. The appropriate equation obtained from (13) by allowing the two relaxation times, T_p and T_s , to vanish. From Part II we know that the J 's contain the respective times as factors, so that they also vanish in this limit. We then have in place of (13)

$$D^2(y - 1) = (K + 4\mu/3)(y - 1) \\ - \eta D(y - 1) - Aq D_t^{-1} D(y - 1) \quad (14)$$

If we operate on this equation by D_t to remove the inverse operator, and use the following notation

$$\left. \begin{aligned} F(y) &= \rho_0 U^2(y - 1) \\ &- (K + 4\mu/3)(y - 1) \\ G(y) &= F(y) - Aqy \end{aligned} \right\} \quad (15)$$

are functions of y alone. Equation (14) can be written

$$d^2 F/dx^2 - Uy dG/dx = (C\lambda d^2/dx^2 \\ - Uy d/dx)(\eta Uy dy/dx) \quad (16)$$

where C is the speed of sound and λ the thermal conductivity path (see Part II for the notation). To integrate this equation, we shall write

$$\left. \begin{aligned} dy/dx, \\ d/dx = (dy/dx) d/dy = z d/dy \\ dy = F', \\ G/dy = G', \quad dz/dy = z \end{aligned} \right\} \quad (17)$$

we can cancel one factor z from (16), which then reads

$$F'' + \eta U^2 y^2 z' + (C\lambda F'' + \eta U^2 y)z \\ - UyG' = UC\lambda\eta (d/dy)(z^2 + yzz') \quad (18)$$

The right-hand side of this equation contains the product of the two physical parameters, λ and η , and it also contains the nonlinear terms in z ; we shall first integrate the equation resulting from a neglect of this expression and then estimate the error. If we write

$$f(y) = (C\lambda F'' + \eta U^2 y)/(C\lambda F' + \eta U^2 y^2) \\ g(y) = UyG'/(C\lambda F' + \eta U^2 y^2) \quad (19)$$

the equation to be solved becomes in first approximation

$$z' + f(y)z = g(y) \quad (20)$$

The solution of this is, provided g is not zero,

$$z = \exp \left[- \int f dy \right] \int g \left[\exp \int f dy \right] dy \quad (21)$$

and then because $z = dy/dx$ we can at once write the formal expression for the profile:

$$x = \int \frac{dy}{\exp \left[- \int f dy \right] \int g \left[\exp \int f dy \right] dy} \quad (22)$$

To check the approximation, we can assume a Hookean solid, in the sense that $K + 4\mu/3$ is independent of y .

The velocity of propagation must be the speed of sound and this depends upon whether the profile is steep enough to be 'isothermal' or whether it is adiabatic. To include both possibilities, we shall write

$$\rho_0 U^2 = K + 4\mu/3 + \gamma Aq \quad (23)$$

where $\gamma = 1$ corresponds to the adiabatic and $\gamma = 0$ to the isothermal limit. We then have

$$\left. \begin{aligned} F &= \gamma Aq(y - 1), \quad F' = \gamma Aq, \quad F'' = 0 \\ G &= (\gamma - 1)Aqy - \gamma Aq, \quad G' = (\gamma - 1)Aq \end{aligned} \right\} \quad (24)$$

Also

$$\left. \begin{aligned} f(y) &= \eta U^2 y / (C\lambda\gamma Aq + \eta U^2 y^2) \\ g(y) &= yU(\gamma - 1)Aq / (C\lambda\gamma Aq + \eta U^2 y^2) \\ \text{with} \\ b^2 &= C\lambda Aq / \eta U^2 \end{aligned} \right\} \quad (25)$$

For the isothermal limit, $\gamma = 0$, one can easily write a profile solution and check whether the width is small compared with the value of λ , the free path for thermal diffusion. In fact, it turns out that the condition for this to be so is in conflict with the condition that the nonlinear terms on the right side of (18) be small. The method therefore breaks down in this case.

We return instead to the adiabatic case, $\gamma = 1$. Here obviously $g(y)$ is zero, and the integral in (21) is of no use. We go back to (20) with $g = 0$ and integrate it immediately to give

$$dx/dy = B(y^2 + b^2)^{1/2} \quad (26)$$

where B is an arbitrary constant of integration. A second elementary integration gives the profile in terms of two arbitrary constants, namely, the final compression ratio y_f and the total width of the profile x_f . Making use of the adiabatic condition in the form

$$b^2/y_f^2 \ll 1 \quad (27)$$

and including only terms of first order in this small ratio, we have

$$x/x_f = \frac{y_0^2 - y^2 + b^2 \ln(y_0/y)}{y_0^2 - y_f^2 + b^2 \ln(y_0/y_f)} \quad (28)$$

This shape is approximately parabolic; it has a finite gradient at both ends of the profile, so that a discontinuity in slope occurs both at the onset and at the finish. There is no relation between x_f , the width, and $y_f - y_0$, the depth or amplitude of the profile, both of which are, of course, arbitrary (Fig. 9). For the result to be valid, the width has to be large enough to make sure that $\gamma = 1$, and this means that (27) must be true, or, in more explicit notation,

$$\lambda A q / \eta U \ll y_f^2 \quad (29)$$

Putting reasonable numerical values in (29), we find that for solids with metallic conductivities the viscosity needs to be not less than 10^{-3} cgs poises for the approximation to be valid. For solids with low conductivities, like silica or quartz, it is necessary for the viscosity to be much greater, by factors of a thousand or a hundred, respectively. The profiles would probably be realistic for fused rock. The scale factor or width of the profile is of the order of magnitude of millimeters or centimeters; it is therefore

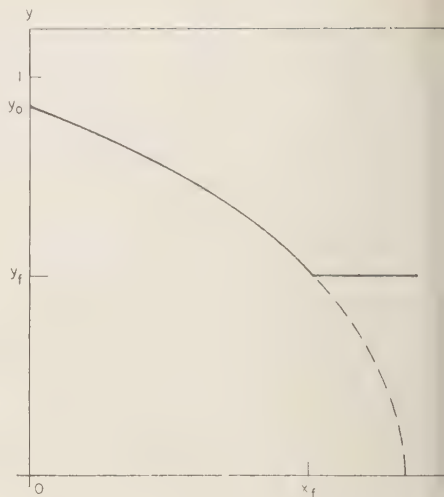


Fig. 9. Profile of equation 28, Part III.

not even remotely like a shock. Since the solid has been assumed Hookean, the Hugoniot jump conditions require that U , the velocity of propagation of the profile, be equal to C , the speed of sound, and that the discontinuities in slope at the onset and finish of the profile travel with the same speed. The 'permanence' of the profile can be rationalized physically as follows.

The thermal conductivity tends to spread the profile: the greater the heat flow, the greater the deviation from 'adiabatic' conditions; the slower the propagation speed; therefore, the steeper part of the profile tends to be retarded and the initial foot of the profile tends to move faster. The viscosity, on the other hand, tends to increase the speed of propagation of the steeper part of the profile through its dispersive effects and to retard the initial change in slope at the foot of the profile. These two effects balance each other in the profile of the permanent regime.

The completely unrelaxed case. Here we suppose that the time of duration of the profile at any location in the material is so short that the relaxation process makes negligible progress until well after the shock has passed. In the permanent regime itself the situation is then completely unrelaxed, and the appropriate equation is obtained from (13) by making the results of the operation $T_d D$, $T_s D$ much larger than unity. In other

ds, we may neglect the unity in each of the relevant factors. Dividing through by $T_p T_s$, we arrive at the following:

$$\begin{aligned} D^2(y-1) &= D^2(K + 4\mu/3)(y-1) \\ &+ (J_p/T_p + 2J_s/3T_s) D^2(y-1) \\ &+ \eta D^3(y-1) \\ &- Aq D^2 D_t^{-1} D(y-1) \end{aligned} \quad (30)$$

sufficiently general relation is obtained from by equating the operand of the D^2 operator to —which simply means that we can cancel D^2 in the equation. The result is an equation initially the same as (14) except that the term $+4\mu/3$ is replaced by $K + 4\mu/3 + J_p/T_p + 2J_s/3T_s$. Hence the only effect of the absence of relaxation is to increase the velocity of propagation without altering the form of the profile.

Partially relaxed cases. To decide whether relaxed or unrelaxed equation is appropriate in any particular example, one has to estimate the time of duration of the profile. Thus, if the x_f/U in (28) is comparable to either T_s or neither extreme outlined above can be safely used. The intermediate case is extremely complicated, and in order to simplify the problem as a first approach we shall, as before, suppose the material to be Hookean, so that $K + 4\mu/3$ is independent of y , and also that the relaxation mechanism is coupled only to the shear strain, $\epsilon = 0$. (Alternatively, we could, of course, couple it only to the compressive strain, $J_s = 0$.) The result is that the operator, $1 + T_p D$, is a common prefactor on the equation of motion and can be canceled from the equation without significant loss of generality. To handle the thermal term, we shall first suppose the thermal conductivity to be high and the free path λ to be short enough so that the profile may be such that in the operator D_t the time derivative term is dominant:

$$D_t = -D \quad (31)$$

which effectively restricts us again to the adiabatic limit.

Equation (13) now reduces to

$$\begin{aligned} U^2 - K - 4\mu/3 - Aq \\ \cdot (1 + T_s D)(y-1) \end{aligned}$$

$$= (1 + T_s D) D(y-1) + \frac{2}{3} J_s D(y-1) \quad (32)$$

Because in the adiabatic case the first factor vanishes, the equation reduces to

$$\begin{aligned} D(D + a)(y-1) &= 0, \\ a &= (\frac{2}{3} J_s + \eta)/\eta T_s \end{aligned} \quad (33)$$

This equation integrates by elementary steps to give the profile

$$x = (U/a) \left(y_0 - y + y_f \ln \left[\frac{y_0 - y_f}{y - y_f} \right] \right) \quad (34)$$

To be acceptable, this result must in practice be such that U/a , the 'width,' is long compared with λ , the thermal free path, so that the adiabatic approximation remains valid.

This profile has an entirely different character from that found in the previous section where the relaxation mechanism was inactive. The present profile has vanishing slope at y_f and reaches that level only asymptotically. The profile does not approach the unrelaxed limit form as the relaxation time increases; in (34) such an approach would simply flatten the profile. On the other hand, neither would the profile approach the completely relaxed limit if the relaxation time were to approach zero; in (34) such an approach is actually forbidden because it would make the width of the profile so narrow that eventually the adiabatic approximation from which the deduction started would be invalidated. This discontinuity between the solutions of a differential equation as a parameter approaches some singular value is not an unusual difficulty in this field of study.

THE ISOTHERMAL APPROXIMATION, HOOKEAN SOLID

It is fairly easy to develop a crude approximation to the isothermal (steep profile) limit. Thus, we consider the operator D_t for a very steep profile to be such that the second space-derivative term dominates; this, in turn, means that $|D_t^{-1}D| \ll 1$, effectively, and we may as a crude approximation simply drop off the term containing this product in (13). Since in this case also $\rho_0 U^2 = K + 4\mu/3$, the first two terms of (13) cancel, and it reduces to

$$(1 + T_s D)(1 + T_p D) D\eta(y - 1) + [J_p(1 + T_s D) + (2J_s/3) \cdot (1 + T_p D)] D(y - 1) = 0 \quad (35)$$

In the relaxed limit, the T 's $\rightarrow 0$, and this reduces to

$$(\eta + J) D(y - 1) = 0$$

and the only solution is $y = 1$, or no profile at all. In the completely unrelaxed limit the equation becomes

$$[T_s T_p D^3 \eta + (J_p T_s + \frac{2}{3} J_s T_p) D^2 + \eta(T_s + T_p) D^2](y - 1) = 0 \quad (36)$$

Canceling D^2 and equating the remaining factor to an arbitrary constant, we can integrate this in elementary steps to find

$$x = (U/a)\{y_0 - y + Y_f \ln \cdot [(y_0 - y_f)/(y - y_f)]\} \quad (37)$$

where

$$a = (\eta + J_p)/T_p \eta + (\eta + 2J_s/3)/T_s \eta \quad (38)$$

If we had supposed the relaxation mechanism to be coupled only with the shear strain, so that $J_p = 0$, the factor $(1 + T_p D)$ could have canceled from (35). Replacing (36) there is now a second-order equation in D , which again integrates to precisely the same form as (37) but with the parameter

$$a = (\eta + 2J_s/3)/T_s \eta \quad (39)$$

This result is identical with (33) and (34) for the adiabatic profile in the partially relaxed case. If the relaxation time T_s is comparable to λ/C , the time for thermal diffusion, any profile steep enough to be 'isothermal' would also be unrelaxed, and (37) and (39) would apply. For the same material any profile not steep enough to be isothermal, i.e., approaching adiabatic, would also be partially relaxed, and (33) and (34) would apply. Thus it is conceivable that a continuous family of different permanent-regime profiles can propagate in such a material, depending only on the initial shape applied at the boundary.

We can improve the above results by the method of successive approximations. We have,

in (35), simply dropped the thermal term. We now reinstate that term and at the same time simplify matters by taking $J_p = 0$; the correct equation then reads

$$(1 + T_s D) D(y - 1) + (2J_s/3) D(y - 1) = Aq(1 + T_s D) D_i^{-1} D(y - 1) \quad (40)$$

We shall treat the right-hand side as a perturbation. The solution in the unrelaxed case for the unperturbed equation is, of course, (37) with (39). We use this unperturbed solution in the perturbing term and express this term as a function of y . From (37) we have

$$Uy dy/dx = -a(y - y_f) = Dy \quad (41)$$

We let

$$D_i^{-1} D(y - 1) = D_i^{-1} Dy = \phi(y) \quad (42)$$

so that

$$Dy = D_i \phi(y)$$

which is equivalent to

$$-a(y - y_f) = C\lambda(dy/dx)(d\phi/dy) \cdot [(dy/dx)(d\phi/dy)] - Uy(dy/dx)(d\phi/dy)$$

Simple manipulation reduces this equation where we use (41) and $C = U$:

$$(d/dy)[(1 - y_f/y) d\phi/dy] + (U/a\lambda)y(1 + d\phi/dy) = 0 \quad (43)$$

First, we shall suppose $d\phi/dy \ll 1$ and drop the term from the last part of (43), so that integration can proceed at once to yield

$$d\phi/dy = -(U/2a\lambda)y(y + y_f) \quad (44)$$

and therefore

$$\phi = -(U/12a\lambda)(2y^3 + 3y_f y^2 - 5y_f^3) \quad (45)$$

We then use (44) in the neglected term $d\phi/dy$ (43) to express that part of the equation as a function of y ; the equation again integrates immediately to furnish the second approximation to $d\phi/dy$ as a function of y :

$$d\phi/dy = -(U/2a\lambda)[y/(y - y_f)]\{(y^2 - y_f^2) - (U/4a\lambda)[y^3(y - y_f) + (7/3)y_f(y^3 - y_f^3)]\}$$

$$D\phi = (U/2\lambda)y\{y^2 - y_f^2 - (U/4a\lambda)[y^3(y - y_f) + (7/3)y_f(y^3 - y_f^3)]\} \quad (46)$$

Multiplying this last expression by AqT_s gives us precisely the perturbation required for (40) in the unrelaxed case under discussion. The perturbed equation is now

$$T_s D^2y + (2J_s/3) Dy = AqT_s D\phi(y) \quad (47)$$

We can cancel one D factor and write instead

$$T_s D(y - y_f) = -(2J_s/3)(y - y_f) + AqT_s[\phi(y) - \phi(y_f)] \quad (48)$$

where $\phi(y)$ can be found from (46):

$$\begin{aligned} \phi(y) - \phi(y_f) = & -(U/12a\lambda)(2y^3 \\ & + 3y_fy^2 - 5y_f^3) \\ & + (1/10)(U/12a\lambda)^2[36y^5 + 7y_f(15y^4 \\ & + 20y_fy^3 + 30y_f^2y^2) - 491y_f^5] \end{aligned} \quad (49)$$

Equation (48) can also be written in the form

$$= - \int_{y_f}^y \frac{U\eta T_s y dy}{(2J_s/3)(y - y_f) + AqT_s[\phi(y_f) - \phi(y)]} \quad (50)$$

which is, of course, the profile. While the algebra would undoubtedly be tedious, it is possible in principle to integrate the expression in (50) by elementary methods, using the particular form of function given in (49) for $\phi(y)$. The width or scale factor is the same as that for the profile (37), but the perturbation due to thermal terms modifies the detailed shape of the profile.

$$x = - \int_{y_f}^y \frac{U_f y dy}{\rho_0(U_f^2 - U_v^2)(1 - y)} + \frac{U_f y dy}{(2J_s/3T_s)(y - y_f) - Aq(\phi_v - \phi_f)} \quad (56)$$

THE SHOCK PROFILE IN NON-HOOKEAN SOLID

All the profiles obtained up to this point were derived on the assumption that Hooke's law was valid; i.e., the elastic modulus $K + 4\mu/3$ as a constant, independent of the compression. Because of this, none of the profiles can be applied directly to the propagation of true shocks. Having now developed the equations

for steady-state profiles in unrelaxed near-isothermal conditions, it should be possible for us to discuss true shocks. To do so we must allow the elastic moduli to be dependent on y . We shall start from (13) and incorporate the various simplifications adopted in the last section, namely, $J_p = 0$, near-isothermal and unrelaxed conditions, which are in fact appropriate for the very steep profiles of shocks; but we shall allow the elastic modulus to be some unspecified function of y .

In (13) the first two terms no longer cancel. Instead we write

$$p(y) = (K + 4\mu/3)(1 - y) \quad (51)$$

and allow this to be some unspecified function of y . In fact, we shall use the Hugoniot jump conditions and the same reasoning as explained in Part I to express p in terms of the shock propagation speed that corresponds to p as a final longitudinal pressure:

$$-p(y) = \rho_0 U_v^2 (y - 1) \quad (52)$$

where U_v is a function of y ; the U that has already appeared in (13) is the shock speed $U = U_f$ given by the final pressure in the profile:

$$p_f = \rho_0 U_f^2 (1 - y_f) \quad (53)$$

The equation replacing (47) is now

$$\begin{aligned} T_s D[\rho_0(U_f^2 - U_v^2)(y - 1)] \\ = \eta T_s D^2(y - 1) \\ + (2J_s/3) D(y - 1) - AqT_s D\phi \end{aligned} \quad (54)$$

Again, we can cancel the operator D and obtain

$$\begin{aligned} \eta T_s D(y - y_f) = \rho_0 T_s (U_f^2 - U_f^2)(y - 1) \\ - (2J_s/3)(y - y_f) + AqT_s(\phi_v - \phi_f) \end{aligned} \quad (55)$$

This integrates at once to give the profile in the form

$$x = - \int_{y_f}^y \frac{U_f y dy}{\rho_0(U_f^2 - U_v^2)(1 - y)} + \frac{U_f y dy}{(2J_s/3T_s)(y - y_f) - Aq(\phi_v - \phi_f)} \quad (56)$$

This expression contains the same terms that were found in Part I, equation (25), and, in addition, the perturbations due to relaxations and thermal effects. It is evident that numerical methods of integration will be needed in order to see the exact form of the profile in any particular case.

In Part I we showed that the total stress versus volume is a straight line through the

TABLE 1. Summary of the Various Cases Studied in Part III

Case	Profile Form	Propagation Speed
Hookean solids		
Adiabatic		
Fully relaxed	Eq. (28)	Normal speed of sound
Completely unrelaxed	Eq. (28)	Increased speed of sound
Partially relaxed	Eq. (34)	Intermediate speed of sound
Isothermal		
Fully relaxed	No solution possible	
Completely unrelaxed	Eq. (37)	Isothermal speed of sound
Partially relaxed	Eq. (50)	Isothermal speed of sound
Non-Hookean solids		
Isothermal unrelaxed	Eq. (56)	Speed of shock wave from Hugoniot jump conditions

profile. This, in fact, can be shown to be a thermodynamic necessity from the 'permanence' of the profile. We can see now that, even with the thermal corrections, the same result holds. Thus, from (8) with $J_p = 0$, the total stress is modified by viscous and thermal terms as well as by relaxation terms. In detail we have

$$p^* = p - \eta y U_f dy/dx + Aq(\phi_u - \phi_f) - (2J_s/3T_s)(y - y_f) \quad (57)$$

Substituting this into (55) at once gives

$$p^* = p_f - \rho_0 U_f^2 (y - y_f) \quad (58)$$

which, with the Hugoniot jump conditions, results in the linear relation once again:

$$p^* = \rho_0 U_f^2 (1 - y) \quad (59)$$

HEATING EFFECT OF SHOCKS

In the theory so far we have neglected the irreversible heat produced by the viscous forces during the yielding process. Such a term ought to be taken into account in the conservation of heat, equation (3b), and in principle it is not difficult to do this, although none of the equations so obtained can be solved exactly. Thus the corrected equation for temperature, replacing (3b), is

$$D_t \theta = q de/dt - (\eta/\rho_0^2 C_s)(de/dt)^2 \quad (60)$$

The new term here is the rate at which heat is being produced by work done against viscosity, divided by the heat capacity. It has to be

treated as a perturbation because it is nonlinear. If we restrict ourselves to steep profiles in which the 'isothermal' limit is a good first approximation, the perturbation method is satisfactory; in fact, the perturbing term in the right side of (40) is simply modified to read

$$Aq(1 + T_s D) D_t^{-1} [Dy - \mu(Dy)^2] \quad (61)$$

where

$$\mu = \eta/\rho_0^2 q C_s \quad (62)$$

Equation (42) for ϕ is now replaced by

$$Dy - \mu(Dy)^2 = D_t \phi \quad (63)$$

and again, if (41) is used for Dy , this reduces to a simple algebraic equation for $d\phi/dy$:

$$d\phi/dy = (U/a\lambda)y[\frac{1}{2}(y + y_f)(1 + ay_f) - (a\mu/3)(y^2 + yy_f + y_f^2)] \quad (64)$$

This integrates in a straightforward fashion to give the modified function ϕ . In this way we may proceed exactly as in (46) to (50), obtaining formally the same result, except that the modified form of ϕ has to be used in the profile integral (50) to take care of the heating effect. A similar argument puts the modified ϕ into (56) for the shock profile. Again, numerical methods are needed in order to see the detailed meaning of the equation in any particular case.

Of greater interest is the temperature in the profile. This can be derived in principle from (60) combined with the profile, i.e., with y as a function of x . As a first approximation we may

the unperturbed profile of (56) and the corresponding expression for the derivative applied in (55); thus we have $de/dt = dy/dt = Dy$, and (55) gives Dy as a function of y . Now (60) can be written in the form

$$C(y) \frac{Dy}{dy} \frac{d}{dy} [(1/Uy) Dy (d\theta/dy)] - Dy = q Dy(1 - \mu Dy) \quad (65)$$

which can be expressed in the form

$$A(y) d^2\theta/dy^2 + B(y) d\theta/dy + C(y) = 0 \quad (66)$$

with all coefficients simple functions of y . This equation can be integrated by elementary means to give θ as a function of y , which, from the profile, can be translated into a function of x . Again, little insight can be obtained without using numerical methods.

DISCUSSION

Experimental data on shock profiles in solids are practically nonexistent. Some data do exist on the heating effects of shocks, and it is planned to program some of the formulas for computer work. Meanwhile, the general results given in this paper would seem to be of sufficient interest to themselves.

It has been repeatedly emphasized that permanent-regime or steady-state profiles should not be confused with actual profiles from explosive-induced shocks. However, it should also be remarked that the Hugoniot jump conditions, which are always used in the calculations of peak pressures in explosive-induced shocks, are also true only in the permanent-regime case. Therefore, the profiles that we calculate by this method are, in principle, just as good as the peak pressure values conventionally accepted.

We would suggest that actual shocks may extend towards the permanent-regime profiles with propagation through attenuating material, and that the next step in theory must be to attempt a perturbation treatment of the non-linear equations for the difference between actual and permanent-regime profiles.

Acknowledgments. The writer is gratefully indebted to D. R. Curran and D. G. Doran who have done all the numerical computations involved in this paper, including the setting up of a program on the IBM 650 computer at Washington State University, to evaluate the integral of equation (25). He also wishes to thank Dr. George E. Duvall

for introducing him to this field of study and for many valuable discussions.

Publication supported by the National Science Foundation. Work partially performed under USAF Contract No. AF-49-(638)-625.

REFERENCES

- Al'tshuler, L. V., K. K. Krupnikov, and M. I. Brazhnik, Dynamic compressibility of metals under pressures from four hundred thousand to four million atmospheres, *Soviet Phys. JETP*, **34**, 614-619, 1958.
- Band, William, *Introduction to Mathematical Physics*, D. Van Nostrand, Princeton, 1959.
- Bancroft, D., E. L. Peterson, and S. Minshall, Polymorphism of iron at high pressures, *J. Appl. Phys.*, **27**, 291-298, 1956.
- Born, M., and K. Huang, *Dynamical Theory of Crystal Lattices*, Univ. Press, Oxford, 420 pp., 1954.
- Cole, R. H., *Underwater Explosions*, Princeton Univ. Press, 1948.
- Doran, D. G., G. R. Fowles, and G. A. Peterson, Shock wave compression of aluminum, *Phys. Rev. Letters*, **1**, 402-404, 1958.
- Drummond, W. E., Shock and detonation theory, *Stanford Research Inst., Internal Rept. 012-58*, 55 pp., 1958.
- Duvall, G. E., and J. S. Koehler, Use of hydrodynamics in treating shock waves in solids, *Bull. Phys. Soc.*, **4**, 283, April 30, 1959.
- Gilbarg, D., The existence and limit behavior of the one-dimensional shock layer, *Am. J. Math.*, **73**, 256-274, 1951.
- Hirschfelder, J. O., C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, John Wiley & Sons, New York, 1954.
- Jaeger, J. C., *Elasticity, Fracture and Flow*, Methuen, London, 1956.
- Mason, W. P., *Piezoelectric Crystals and Their Applications to Ultrasonics*, D. Van Nostrand, Princeton, 508 pp., 1950.
- Minshall, S., Properties of elastic and plastic waves determined by pin contactors and crystals *J. Appl. Phys.*, **26**, 463-469, 1955.
- Morse, P. M., and H. Feshbach, *Methods of Theoretical Physics*, McGraw Hill, New York, 1953.
- Rayleigh, Lord, Aerial plane waves of finite amplitude, *Proc. Roy. Soc. London, A*, **84**, 247, 1910.
- Rice, M. H., R. G. McQueen, and J. J. Walsh, Compression of solids by strong shock waves, *Solid State Phys.*, **6**, p. 1, 1958.
- Walsh, J. M., M. H. Rice, R. G. McQueen, and F. L. Yarger, Shock wave compression of twenty-seven metals. Equation of state of metals, *Phys. Rev.*, **108**, 196-216, 1957.
- Zener, C., *Elasticity and Anelasticity of Metals*, Univ. Chicago Press, 170 pp., 1958.

(Part I received July 20, 1959; Part II received August 7, 1959; Part III received August 29, 1959.)

Seismograms Associated with the Near Passage of Tornadoes

CARL KISSLINGER

*St. Louis University
St. Louis, Missouri*

Abstract. The principal features found on seismograms obtained on three occasions when a tornado passed near the station are described. Long periods, 50 sec and greater, predominate, with periods of 30 sec and less and bursts of very short period activity superimposed. The classical solution for a vertical load on an elastic half-space is applied, and the results indicate that elastic tilts due to the static loading effect of the tornado contribute substantially to the basic form of the record. The shorter periods are ascribed to the dynamic loading of the high winds on obstacles in the path. Part of the activity is not ground motion, but is due to the buoyancy effect caused by the rapid pressure fluctuations.

Introduction. In view of the current interest in the problem of ground motion produced by a pressure pulse moving through the air over the earth's surface, it was thought worthwhile to examine the records that have been obtained from St. Louis University seismographs during the close passage of tornadoes. Considering that these seismograph stations are located within the 'tornado alley' of the mid-continent, one is not surprised to find that on a number of occasions these very severe local storms have passed at short distances from the vaults. For each one of these occurrences for which the records have been examined, activity generated by the tornado has been recorded. It is the purpose of this paper to describe the recorded events of a few of these storms and to speculate as to the predominant mechanism that produces ground motion.

The first problem one is faced with in examining the records is the uncertainty as to whether the motion recorded is all truly ground motion or is at least partly due to the direct effect on the pendulums of the large, rapid fluctuations in atmospheric pressure that accompany a tornado. For example, there is no doubt that long-period vertical seismographs respond to atmospheric pressure oscillations unless they are carefully compensated for such variations [Ewing and Press, 1953]. The evidence available does confirm, however, a belief that some of the records do represent genuine ground motion.

The records from the 80-kg Wiechert horizontal seismograph are of value in determining whether ground motion is present. This instrument was operating at St. Louis University in 1927 when the disastrous tornado of September 29 passed within about half a mile of the station. The Wiechert vault is a small, completely unvented ground floor room equipped with a massive door. Though the door is not intentionally sealed against transmission of pressure variations, its construction is such that it would not seem likely to pass such fluctuations readily. As will be seen, a small pressure effect is present on these records, but it can be distinguished from what is thought to be the ground motion. Another great advantage of this purely mechanical instrument is that it records slow tilts.

The Wilip-Galitzin instruments at Florissant, Missouri, are enclosed in the usual glass covers with tight grease seals, and seem to be relatively insensitive to air pressure variations of short duration. The new long-period Sprengnether vertical seismograph at the gymnasium station is enclosed in a sealed rigid metal cover, but it appears that this cover buckled slightly under the most severe rapid pressure changes accompanying the St. Louis tornado of February 10, 1959. The horizontal Sprengnethers are under plastic covers, with no provision for sealing.

Figure 1 shows the tracks of the tornadoes of 1927 and 1959 through St. Louis, and the loca-

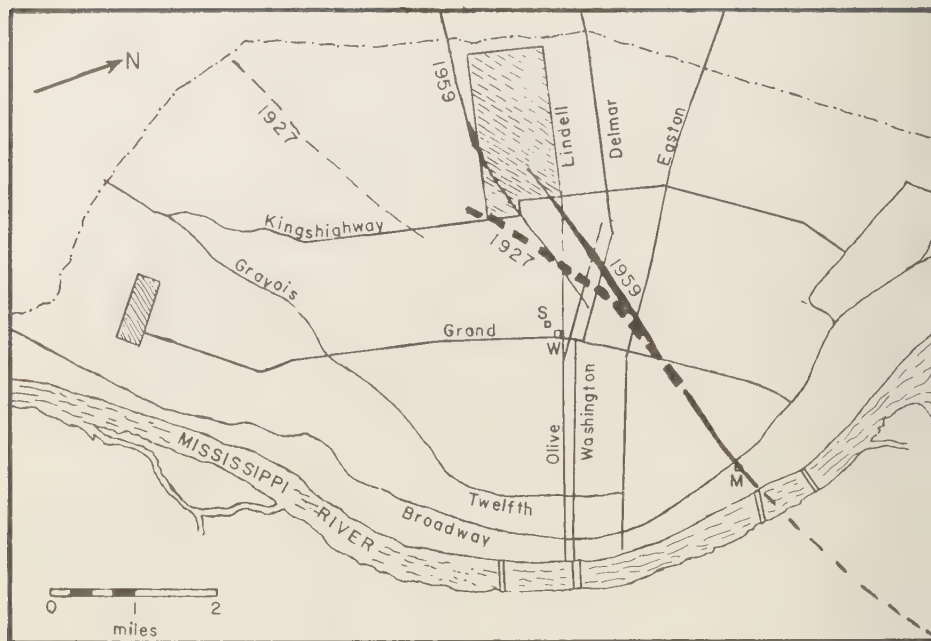


Fig. 1. Tornado paths, 1927 and 1959, in St. Louis, Missouri.

tions of the seismograph installations. The width of the track indicates the severity of the tornado on the basis of damage. *W* indicates the Wiechert location, at which the 1927 storm was recorded, and *S* the gymnasium vault, at which the records of the 1959 storm were obtained. *M* indicates the Mallinckrodt Chemical Company, at which a microbarograph was located within a block of the storm center. It will be noted that the two tracks, from southwest to northeast, are very similar, especially in the vicinity of the stations. In each case there are two parallel tracks, probably representing two separate tornadoes, one dissipating and a new one forming.

The other tornado to be discussed passed about one-fourth of a mile northwest of the Florissant vault on May 9, 1957.

Description of the storms and seismograms. A detailed description of the storms will not be given here; only that information which is helpful in interpreting what is seen on the seismograms will be presented.

The tornado of September 29, 1927, by far

the most severe of the three under discussion occurred at 1300 CST.¹ The high wind velocity of 70 mph from the south lasted from 1258 to 1303 at the U. S. Weather Bureau office in downtown St. Louis, about 2 miles east of the university and about the same distance from the center of the storm path. The maximum wind speed of 96 mph was recorded at 1302. At this time the barograph recorded an almost instantaneous drop of 0.11 inch and then rose immediately 0.22 inch. At the university weather station the drop was 0.2 inch followed by an immediate rise to the previous reading. The width of the path was 300 to 5,280 ft. The tornado started in extreme southwest St. Louis doing sporadic damage. Then either a new storm originated or the original one veered sharply to a new path and intensified in the middle of the city, doing great damage. One eyewitness in midtown estimated the funnel diameter as 600 feet [Hayes, 1927].

The Wiechert record at the time of t

¹ All times in this paper are in CST.

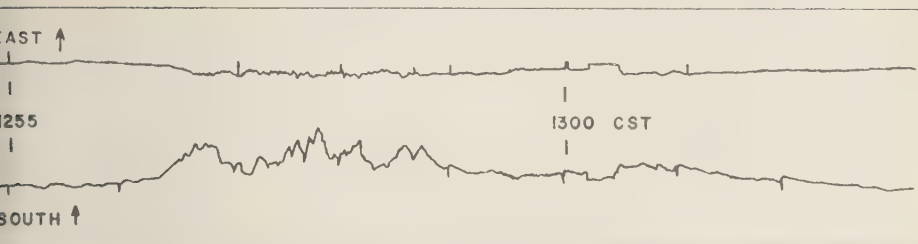


Fig. 2. Seismogram at the time of a tornado; Wiechert horizontal seismograph; September 29, 1927.

nado is shown in Figure 2. Both components had a free period of 6.1 sec and a static magnification of about 80, and both were damped out 0.55 critical. The time marks are at main intervals. The station log shows that both instruments had the same friction constant only a month before, and a strong earthquake was recorded equally well on both components only a few weeks after the tornado. Therefore, we must conclude that the marked difference in activity on the two components indicates a real difference in excitation. This difference is not easily explained, but would seem to confirm the hypothesis that we are dealing with ground motion, since the effect of pressure variations would be nondirectional. The motion can be analyzed as a combination of a slowly rising step of long duration, a long-period oscillatory motion, and bursts of shorter-period oscillations. The long-period activity is much smaller on the east-west record than on the north-south. If all the recorded activity proceeded from the storm center, then both components should show activity with about the same amplitudes, as the direction from the station to the storm changed continuously. If the pronounced oscillatory portion is produced locally at the station by the strong winds, then a correlation between the prominent north-south motion and the strong north winds is seen.

The step occurs simultaneously, down on the east-west and up on the north-south. The duration of the step is 7 min, 37 sec, with the return to the original position somewhat more gradual than the onset. The step is preceded on the north-south by 1 min, 25 sec of small, high-frequency motion. A few cycles of a very long period, 50 to 70 sec, are seen on the north-south

—perhaps the most striking feature on the record. Short-period motion, 10 sec and less, begins at the same time as the long periods, and reaches its greatest intensity from 1257 to 1259, coinciding with the time of intense activity of the storm.

The tornado of February 10, 1959, struck the midtown region of St. Louis at about 0213 CST [*Brancato*, 1959]. Figure 3 shows the records from both long- and short-period seismographs in the gymnasium station. The time of the most pronounced activity on both the vertical and the north-south long-period instruments coincides with the time of the storm. Activity first appears on the long-period vertical at 02^h 08^m 19^s, with a low-amplitude, long-period form. A more pronounced motion with a long period (93 sec) begins at 02^h 10^m 15^s. Two cycles of about 40-sec period are then followed by the most striking feature, a sharp break down (ground up) at 02^h 12^m 50^s, followed by a deflection in the opposite direction at 02^h 13^m 41^s with a period of 48 sec. The motion then trails off in a series of small events of about 30-sec period. It is only during the maximum vertical activity that the long-period horizontal shows any significant motion, beginning at 02^h 12^m 23^s and reaching a small maximum at 02^h 13^m 25^s. The period of the horizontal motion is about 33 sec. The shorter-period motion on these two instruments during the storm looks much like the general background through the day.

On the preceding evening a severe thunderstorm struck the city at about 2000. This was accompanied by a disturbed barograph trace (Fig. 5). The long-period vertical seismograph was disturbed for about an hour, beginning at 1957. Long-period disturbances, similar to those

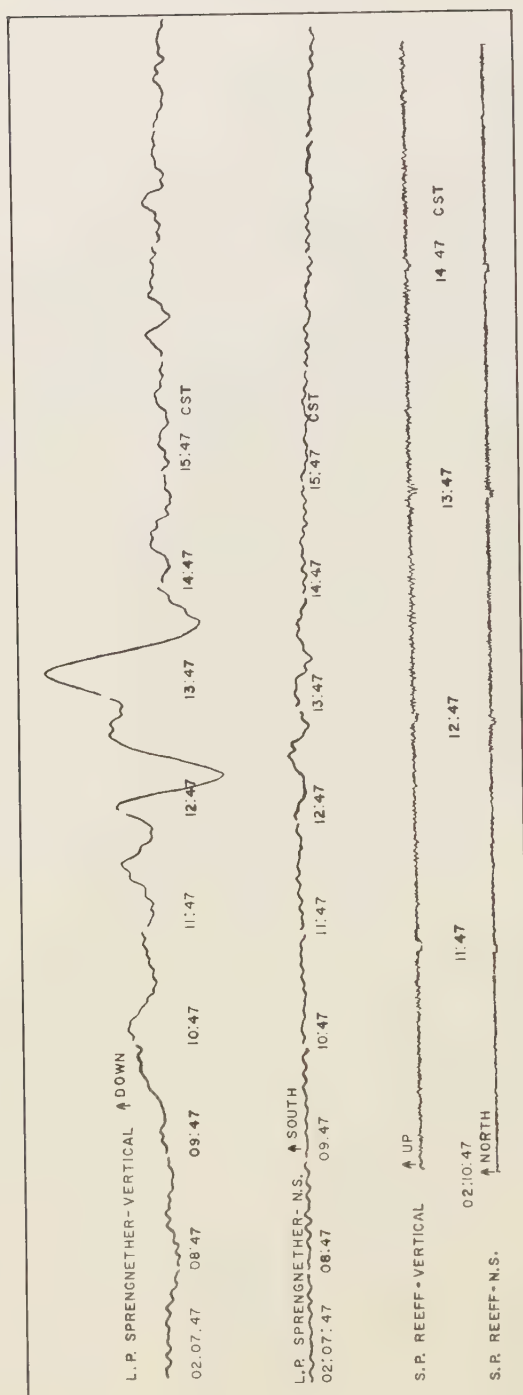


Fig. 3. Seismograms at the time of a tornado; St. Louis station; February 10, 1959.

een in the first three minutes of the record reproduced here, and equally as large, are found. The long-period horizontal was not disturbed during this thunderstorm. This indicates that, even with the metal cover, the vertical was responding to short-period (relative to atmospheric time scale) air pressure fluctuations, while the horizontal was not.

To further test this point, records were examined before, during, and after periods for which the barograph trace had an appearance similar to that at the time of the tornado, without the actual occurrence of a tornado. In each case, the vertical was disturbed during the time of rapid pressure change and became quiet as the barometer leveled off.

It is postulated that the activity on the horizontal component from 0213 to 0215 is truly ground motion, and coincides with the time when the tornado was in midtown St. Louis. The nature of the vertical component is open to question. In view of the abrupt changes in

pressure that occur as a tornado moves past a point, at least part of the activity on the vertical instrument is a buoyancy effect. The sense of the large deflections on the record is correct for a sudden expansion of the case followed by a contraction.

The activity on the short-period Reeff instruments began at the same time as the motion on the north-south long-period instrument. This is taken as further evidence that the ground motion started at this time. The short-period motion, $\frac{3}{4}$ sec to 1 sec, was of a low level and occurred in brief bursts. There is no indication of any long period impressed on the short-period vertical. The duration of the short-period activity was about the same as the principal disturbance on the long-period instruments.

The final case to be considered is that of a small tornado that touched the ground near the Florissant station at about 1445 CST on May 9, 1957 [Parker and Griffith, 1957]. The fun-

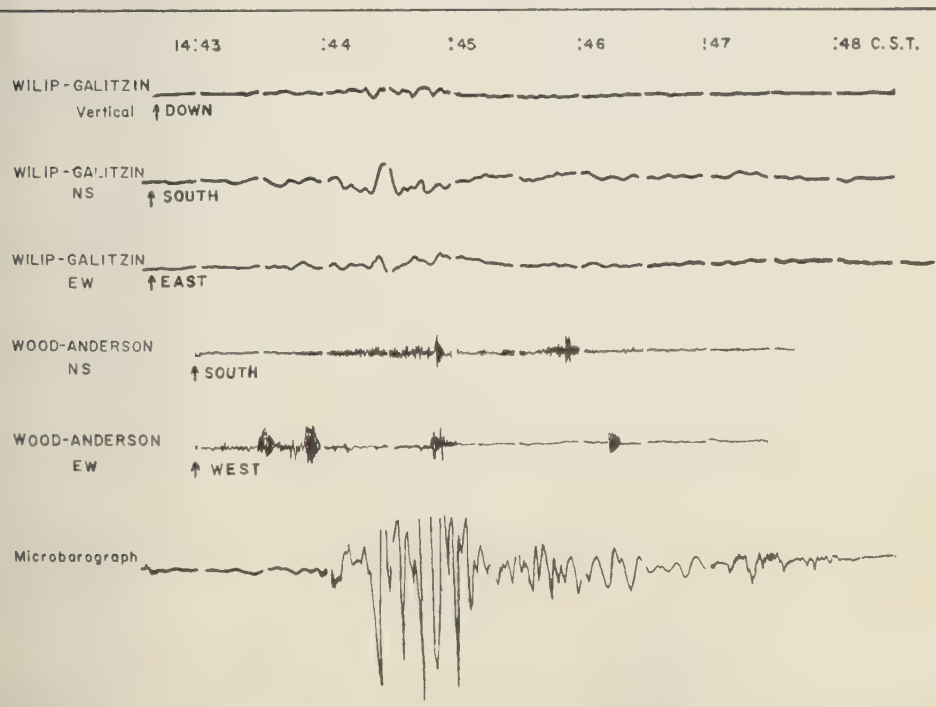


Fig. 4. Seismograms at the time of a tornado; Florissant station; May 9, 1957.

nel came down less than $\frac{1}{4}$ mile to the southwest and passed about 0.1 mile to the west of the station on its northeasterly path. The diameter of the funnel did not exceed 75 feet, with a path of damage about 300 feet in width. Several trees near the station were uprooted.

Figure 4 shows the Wilip-Galitzin and Wood-Anderson records, as well as the trace from a Maceiwane microbarograph. It is seen at once that the greatest activity on all but the east-west Wood-Anderson occurred from 1444 to 1446, in agreement with the observed time of the tornado.

The records of all three Wilip-Galitzin instruments show a combination of a long period (60 sec to 70 sec) with shorter periods superimposed. The peaks of the long period occur simultaneously on the vertical and north-south at $14^h 44^m 30^s$, whereas on the east-west they occur at 45^m . If this activity represents ground displacement, the motion is up, north and east. If the results of elastic tilt have been recorded, the maximum tilt would be down to the south and then to the west as the tornado moved past the station.

The north-south Wood-Anderson shows a train of 1 to 1.5 sec periods, beginning at $14^h 43^m 45^s$, with amplitudes gradually increasing during the time of pronounced long-period motion. The east-west trace is not so active. There are a number of bursts of extremely high frequency on both short-period records. It has been suggested that these may be due to thunder or wind gusts.

Mechanisms by which tornadoes may generate ground motion. As the tornado sweeps over the earth's surface, we may expect it to act as a moving source of both normal and tangential stresses. First, because the tornado has a core of low pressure relative to the atmosphere around it, we expect a static loading effect, or the effect of what might be considered an 'underpressure' (as contrasted with the overpressure in an air blast wave). This core is essentially limited to the funnel or vortex.

The very high winds, cyclonic within the vortex, which are characteristic of a tornado will result in dynamic loading. The stresses on the earth's surface from this loading will depend not only on the wind velocity, but also

on the drag coefficient of the surface. In open, flat country we might expect this to be very small, but as the storm moves through a heavily wooded area or through a city complex, we might anticipate large stresses in the earth as obstacles are encountered. In addition, the high winds around the tornado, outside the vortex will create ground disturbances over a wide area.

In general, low-frequency disturbances are expected from the static effect, and disturbances of higher frequency, with a somewhat random, impulsive nature, should result from the dynamic loading. Even if a theory for the motion produced by a moving point source of combined normal and tangential stresses on a layered half-space existed, there would be great diffi-

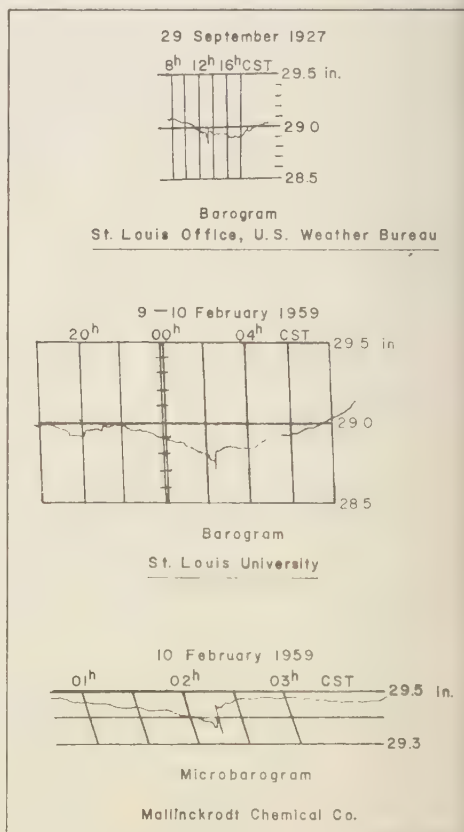


Fig. 5. Barograms at times of tornado passages.

culty in applying it in an attempt to explain the observations quantitatively. This is because a tornado is a very complex and often poorly defined source with which to work. Pressures and wind velocities within the core can be estimated only indirectly; the size of the funnel must be decided from damage and from descriptions by excited witnesses; it is hard to decide when the tornado was on the ground and when it was lifted; and, worse, from the point of view of the seismologist, it is impossible to position the tornado at any particular instant with any kind of exactness. In spite of all of this, it was thought worthwhile to estimate the static loading effect on the basis of an idealized tornado, using approximate theory.

Since the pressure-time history at a point must enter into any mechanism, the available barograph traces for the two St. Louis tornadoes are shown in Figure 5. Certain features are repeated: a step down in pressure a few minutes before the tornado, a levelling off, and then a steep drop followed by an immediate rise, as described earlier.

Since the tornado moves at a speed that is very slow relative to the speed of elastic waves in even a poorly consolidated material, it was thought reasonable [Cole and Huth, 1958] to

compute the ground displacements and ground tilts to be expected from a model tornado by the use of the well-known static solution of Boussinesq for a point load acting vertically on an elastic half-space. Figure 6 shows the results of this effort. Average values for the size and speed of travel of the tornado were taken from the *Compendium of Meteorology*. An underpressure of 1 psi was taken as a reasonable guess, and it may be too low or too high for a given case by a considerable amount. The tornado was pictured as a core of low pressure with a radius of 375 feet.

The medium was taken to be a thick unconsolidated overburden, and layering was neglected. A rigidity modulus of 2×10^4 psi and a value for Poisson's ratio of 0.35 were assumed. These constants correspond to a compressional wave velocity of 1090 ft/sec and a shear wave velocity of 525 ft/sec. Displacements and tilts were calculated, assuming that the tornado moved from southwest to northeast at 35 mph. The point of closest approach to the station was taken as 2500 feet to the northwest. Time marks are shown at 1-min intervals.

The tilts have been converted to the trace displacements they would produce on the

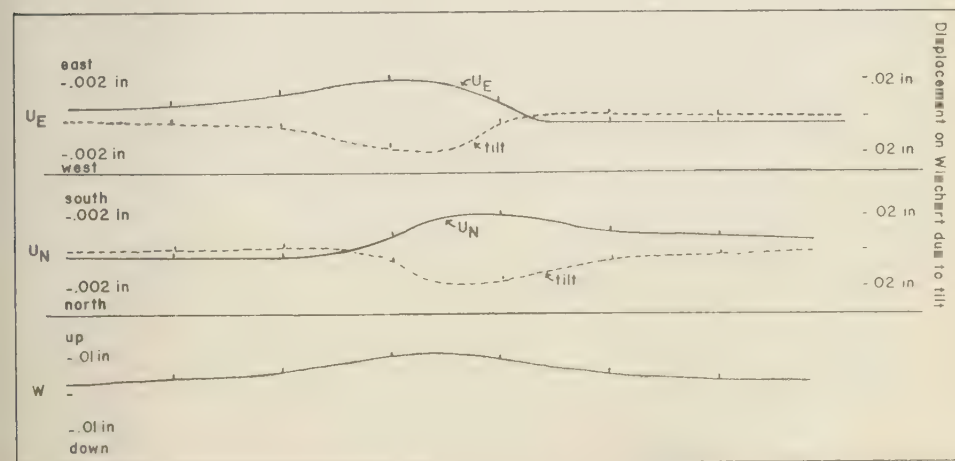


Fig. 6. Ground motion produced by a slowly moving core of low pressure—static approximation. U_E and U_N are the east-west and north-south components of horizontal displacement; W is the vertical displacement. Tilts have been converted into equivalent displacements on the Wiechert horizontal seismograph.

Wiechert, using the sign convention of that instrument. The amplitudes due to tilt are smaller than the observed values by a factor of 5, which is satisfactory when we consider all the simplifications used. The trace form bears some similarity to the over-all appearance of the Wiechert record. However, the sense of the tilt is incorrect for the north-south.

The displacements are plotted for interest only, as the magnification of all the instruments is practically zero for motions as slow as these. It must be taken as a coincidence that if the curve for U_N is multiplied by the static magnification of the Wiechert, an excellent fit to the general form of the north-south record is obtained. The vertical displacement is seen to be greater than the horizontal.

The long step on both components of the Wiechert can be explained, at least in part, by a sudden decrease of air density in the vault due to a pressure drop. On the assumption that the full 0.2-inch drop recorded on the barograph in the same building in 1927 was transmitted to the instrument room, the trace displacement was calculated. The change in air density corresponding to this pressure change was computed on the basis of an adiabatic expansion. The change in the buoyancy force acting on the moving mass of the seismograph was then found, using the known volume of that mass. The distance by which the center of mass of the moving system was displaced from the vertical line through the hinge can be estimated with fair accuracy. With the moment on the system due to the change in buoyancy known, the magnitude and sense of the trace displacement was found. The magnitude was somewhat less than that observed for the step, with the same sense.

In spite of some disagreement in the observed

and theoretical directions, it is postulated that the very long period effects observed are due to elastic tilts resulting from static loading by the tornado. In addition to the basic oscillatory character produced by the continuously changing azimuth, lifting and lowering of the funnel or changes in the storm intensity may also give rise to long-period oscillations. Such effects are unpredictable. The directional characteristics of the Wiechert data indicate that strong winds of almost constant direction on the fringes of the storm contribute to the long-period activity.

The shorter periods, 30 sec and less, are ascribed to dynamic loading, and the very high frequencies are probably the result of wind gusts. It is impossible to decide whether these oscillations are generated within the storm center or near the station by locally high winds, or are, as seems most likely, a combination of both effects.

REFERENCES

- Brancato, G. N., St. Louis tornado, 10 February 1959, *Weatherwise*, 12, 43-44, 1959.
- Cole, J., and J. Huth, Stresses produced in a half plane by moving loads, *J. Appl. Mechanics*, 25, 433-436, 1958.
- Ewing, M., and F. Press, Further study of Atmospheric pressure fluctuations recorded on seismographs, *Trans. Am. Geophys. Union*, 34, 95-100, 1953.
- Hayes, M. W., The St. Louis tornado of September 29, 1927, *Monthly Weather Rev.*, 405-407, September, 1927.
- Parker, W., and W. Griffith, Tornado occurrence and damage, May 9, 1957, unpublished report to Weather Bureau Forecast Center, SELS, Kansas City, Mo.

(Manuscript received October 16, 1959; presented at the Annual Meeting, Eastern Section, Seismological Society of America, June 27, 1959.)

Study of Earthquake Mechanism by a Method of Phase Equalization Applied to Rayleigh and Love Waves¹

KEIITI AKI

*Seismological Laboratory
California Institute of Technology
Pasadena, California*

Abstract. Rayleigh waves and Love waves are used for the study of the earthquake mechanism by the application of a method of phase equalization. In this method, an impulse response is computed from known phase-velocity data and instrument characteristics, and is cross-correlated with an actual record. A comparative study of Love waves from Kern County aftershocks of 1952 and those from Nevada shocks of 1954 strongly supports the hypothesis of a pair of couples rather than a single couple for the earthquake source. Source functions for five Kern County aftershocks are derived from the Rayleigh waves recorded at Weston and Palisades. It was found that the sense of principal motion in the source function is in agreement with the fault-plane solution obtained from the P-wave data. Mantle Rayleigh waves are found to be useful for this purpose also.

Introduction. In a previous paper, Aki [1960] showed that long-period Love waves of continental path can be successfully used for the study of the earthquake mechanism. The method used was simply the comparison between wave forms recorded by the same instrument at the same station due to different earthquakes which were of similar size and which occurred within a limited area. In the present paper, the method was improved so that we can compare the seismograms for different instruments, different epicentral distances, and different wave paths. This method enables us to obtain an equalization of waves with respect to the phase delay due to propagation as well as to recording. We applied this method to Love waves from the Kern County aftershocks of 1952 and the Nevada shocks of 1954, and confirmed the conclusion reported in the previous paper concerning their radiation pattern. We also applied this method to Rayleigh waves and obtained the source functions for Kern County aftershocks and compared them with their fault-plane solutions from the P-wave data.

Method. Analyses and syntheses of seismograms by the Fourier method have been made

by various authors [Sato, 1955, 1956; Valle, 1949]. An ingenious use of a stationary phase technique was devised by Brune and others [1960]. Sato has been interested in getting the impulse form at the source from surface waves for the purpose of studying the earthquake mechanism. In his method, he applies, first, the Fourier analysis to the record and synthesizes the source function from the component waves after correcting the phase factor with respect to the delay due to propagation. We, on the other hand, first compute the impulse response for the wave media and cross-correlate it with the actual seismogram in order to get the source function. Although the two methods are essentially the same, there seem to be considerable differences in the practical computation as well as in the approximation underlying it. With our method more complicated waves and also more weakly dispersed waves can be analyzed than with Brune's method.

We assume that our system, consisting of the dispersive medium plus the instrument, has the following frequency response:

$$G(\omega) = \exp \{-i\phi(\omega)\}$$

$$|G(\omega)| = 1 \quad \text{if } \omega_1 < \omega < \omega_2 \quad (1)$$

$$= 0 \quad \text{otherwise}$$

¹Contribution No. 954, Division of Geological Sciences, California Institute of Technology.

where,

$$\phi(\omega) = \frac{\omega \Delta}{C(\omega)} + \phi_{in}(\omega)$$

ω = angular frequency

Δ = travel distance

$C(\omega)$ = phase velocity

ϕ_{in} = instrumental phase delay.

Here, the amplitude response is approximated by a simple rectangular spectrum.

The impulse response $g(t)$ of our system is expressed as

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty} G(\omega) \exp(i\omega t) d\omega \\ &= 2 \int_{\omega_1}^{\omega_2} \cos[\omega t - \phi(\omega)] d\omega \end{aligned} \quad (3)$$

If we have an impulse response $g(t)$ for a certain wave path and an actual record $x(t)$ for that path, we can get the source function $y(\tau)$ from $x(t)$ and $g(t)$ by the following cross-correlation.

$$y(\tau) = \int g(t)x(t + \tau) dt \quad (4)$$

This equation can be written in terms of Fourier transform as follows.

$$\begin{aligned} Y(\omega) &= G(-\omega)X(\omega) \\ &= G^*(\omega)X(\omega) \\ &= \exp\{i\phi(\omega)\}X(\omega) \text{ if } \omega_1 < \omega < \omega_2 \\ &= 0 \text{ otherwise.} \end{aligned} \quad (5)$$

The cross-correlation, therefore, equalizes the phase angle of the record $x(t)$ to the source in the frequency range from ω_1 to ω_2 . At the same time, this operation works as a low-pass filter to cut off the waves of frequencies higher than ω_2 . The choice of ω_2 is important, especially in the case where the second-mode waves coexist with those of the first mode.

To compute the impulse response $g(t)$, the following approximation was used. Defining a phase-delay time τ by $\tau = \phi(\omega)/\omega$, we rewrite (3) as follows.

$$g(t) = 2 \int_{\omega_1}^{\omega_2} \cos(\omega t - \omega \tau) d\omega \quad (6)$$

Dividing the frequency range from ω_1 to ω_2 into consecutive portions of the interval $\Delta\omega_i$, we again rewrite the equation as

$$g(t) = 2 \sum_i \int_{\omega_i - \Delta\omega_i/2}^{\omega_i + \Delta\omega_i/2} \cos(\omega t - \omega \tau) d\omega \quad (7)$$

To compute the integral for each frequency portion, the argument is expanded into a Taylor series around each ω_i , and the terms higher than $(\omega - \omega_i)$ are neglected. Then a simple calculation gives

$$g(t) = 2 \sum_i \Delta\omega_i \frac{\sin\left\{\frac{\Delta\omega_i}{2}(t - t_i)\right\}}{\frac{\Delta\omega_i}{2}(t - t_i)} \cdot \cos(\omega_i t - \omega_i \tau_i) \quad (8)$$

where t_i is defined as

$$t_i = \tau_i + \omega_i \left(\frac{\partial \tau}{\partial \omega} \right)_i \quad (9)$$

and called the group delay time. It can be shown that this group delay time is equal to the travel time of the wave group of the frequency ω_i which travels with the velocity $U = c - \lambda(dc/d\lambda)$ along the wave path. Equation (8) shows that the component wave for each frequency portion has an envelope of the form $\sin x/x$ which travels with the group velocity, and the amplitude is proportional to the width $\Delta\omega_i$. Although the approximation is better if the frequency range is divided into a greater number of portions, it was found that dividing into more than ten portions is unnecessary for the frequency range of three octaves.

The computation time for one value of $g(t)$ is 20 to 30 sec on the Bendix G-15 electronic computer of the Seismological Laboratory at Pasadena. The cross-correlation computation required for equalization takes a little more time for one value of the source function $y(\tau)$.

Love waves. In the previous paper, it was shown from the comparison of Love waves of a Kern County shock with those of a Nevada shock, both recorded at Palisades, that the hypothesis of a pair of couples is better than that of a single couple. In this comparison, however, we neglected the difference in epi-

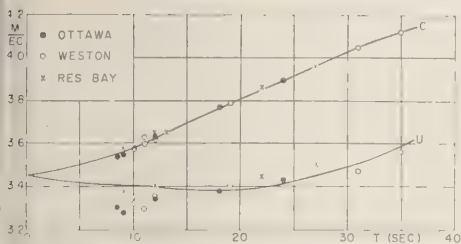


Fig. 1. Phase and group velocities of Love waves obtained from the records of a Kern County aftershock at Ottawa, Weston, and Resolute Bay by Brune's method. The smooth curve is for Dorman's model 208 with the parameters $\beta_1 = 3.45$ km/sec and $H = 38$ km.

central distance between the two shocks because the difference is negligible for the periods longer than 30 sec. In this paper, an equalization with respect to the difference in epicentral distance is applied to seismograms, so that the comparison is possible for shorter periods and also for other stations for which the difference is not negligible even for longer periods.

The records of the Nevada shock of August 31, 1954, and those of the Kern County shock of July 29, 1952, at Palisades and Resolute Bay are used. The fault motions at the sources of both shocks are right-hand strike slips [Båth and Richter, 1958; Aki, 1960]. The fault line lies in the direction N 53° E for the Kern

County shock, and that for the Nevada shock was slightly west of north.

Resolute Bay lies in the quadrant -45° to 45° (with respect to the fault line) for the Nevada shock and in the same quadrant for the Kern County shock (Figs. 4 and 5). On the other hand, Palisades lies in the quadrant 45° to 135° for the Nevada shock and in the quadrant -45° to 45° for the Kern County shock (Figs. 4 and 5). Therefore, if the hypotheses of a single couple is true, Love waves from the two sources must have the same sense at both stations. On the other hand, if the hypotheses of a pair of couples is true, Love waves from the two sources must have the opposite sense at Palisades and the same sense at Resolute Bay.

Equalization of Love waves for nearly common paths. The phase-velocity curve used for the equalization is obtained by the use of Brune's method from the records of a Kern County aftershock at Resolute Bay, Ottawa, and Weston. The experimental points for both phase and group velocities showed little scatter from the theoretical curve of Dorman's model 208 [Dorman, 1959] with β_1 (the shear velocity of the top layer) = 3.45 km/sec and H (the thickness of the crust) = 38 km, (Fig. 1) both of which are reasonable values. Although it is necessary to assume the sense of motion at the source in the application of Brune's method,

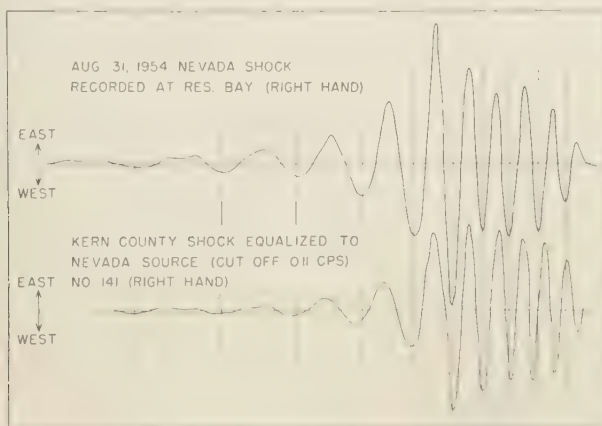


Fig. 2. Comparison of the Resolute Bay record of Love waves from a Nevada shock with that from a Kern County aftershock which is equalized to Nevada source.

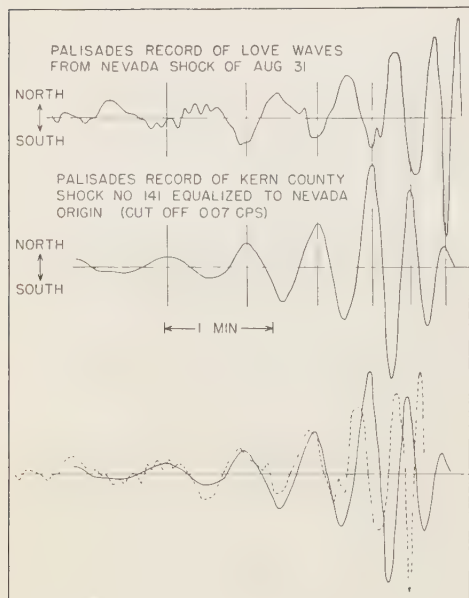


Fig. 3. Comparison of the Palisades record of Love waves from a Nevada shock with that from a Kern County aftershock which is equalized to Nevada source.

the error from a wrong assumption is negligible in this case. This follows because the difference in the epicentral distance between the two shocks is 450 km for Resolute Bay and 190 km for Palisades, and is about equal to or less than 10 per cent of the total epicentral distance. The effect of that error in phase velocity on the wave form equalized from the Kern County source to the Nevada source is therefore very small.

Figure 2 shows the comparison of Resolute Bay records for the two shocks. The Love waves of the Nevada shock were smoothed by a symmetric moving average to eliminate the waves of frequencies higher than 0.11 cps. Since this smoothing has symmetric coefficients, no phase shifts are produced at any frequency. The record compared with this is that of Love waves from Kern County recorded at Resolute Bay and equalized to the Nevada source. The equalization automatically cuts off frequencies higher than 0.11 cps. Both wave forms show an excellent crest-to-crest correspondence for the periods from 50 sec to 12 sec. One cannot get a

good crest-to-trough correspondence, no matter how one arranges the two records by shifting one relative to the other along the time axis. This means that the Love waves from both shocks have the same sense at Resolute Bay.

A similar comparison for Palisades records is shown in Figure 3. Since Love waves from California and Nevada of periods shorter than 20 sec are subject to some disturbances at this station, the comparison was made for the period range from 50 sec to 20 sec. In this case, an excellent crest-to-trough correspondence was observed between the two records. This means that Love waves of both shocks have the opposite sense at Palisades. The lower half of Figure 3 demonstrates that it is impossible to get a good crest-to-crest correspondence between the two records, no matter how one is shifted relative to the other along the time axis.

The above facts are in agreement with the hypothesis of a pair of couples. A recent theoretical study by Knopoff and Gilbert [1959] shows that a displacement dislocation parallel to a fault is equivalent to a pair of couples in the absence of a fault, at least in the radiation patterns for P waves and S waves. It seems therefore, highly probable that an exact computation of radiation pattern in the presence of a fault will also give a quadrant solution for Love waves.

Equalization of Love waves for different paths. So far we have been concerned only with the comparison of waves for which the paths are nearly common. Since the phase-velocity curves for Love waves are not yet available for all regions of North America, it is difficult to compare the wave forms at different stations. In this paper, a tentative study was made using only equalization with respect to the difference in instrumental delay and epicentral distance.

Figure 4 shows the Palisades record of Love waves from the Nevada shock of August 31, 1954, and the Resolute Bay records of the same shock equalized to Palisades with respect to the difference in the instrumental delay and epicentral distance by the use of the phase-velocity curve given in Figure 1. Two records are arranged in two ways in Figure 4: in the upper part, the northward motion at Palisades and the eastward motion at Resolute Bay are taken

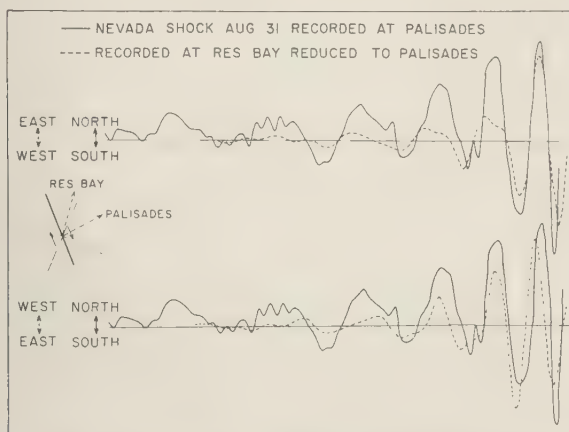


Fig. 4. Comparison of the Palisades record of Love waves from a Nevada shock with the Resolute Bay record of the same waves of the same shock (equalized to Palisades with respect to the difference in the instrumental delay and epicentral distance).

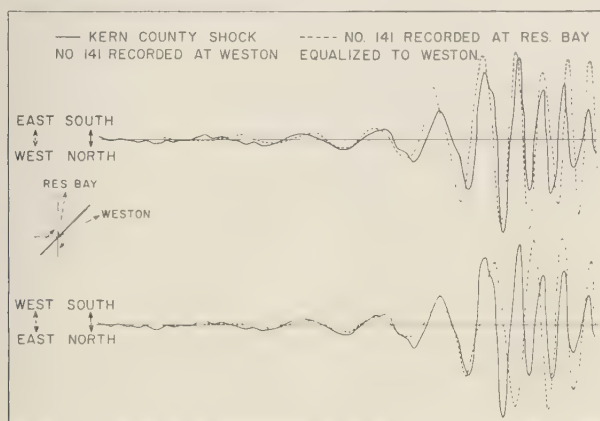


Fig. 5. Comparison of the Weston records of Love waves from a Kern County aftershock with the Resolute Bay record of the same waves of the same shock (equalized to Weston with respect to the difference in the instrumental delay and epicentral distance).

be positive; in the lower part, the sense is reversed for the Resolute Bay record. In both cases the agreement between the wave forms is not good because of poor equalization due to our ignorance about phase velocities for different regions along the wave path. But the agreement seems better in the arrangement shown in the upper part of Figure 4. This indicates that the northward motion at Palisades corresponds to the eastward motion at Resolute

Bay, in agreement with the result obtained in the preceding section.

A similar comparison was made between the Weston records of the Kern County aftershock of July 29, 1952 (No. 141), and the Resolute Bay record of the same earthquake equalized to Weston (Fig. 5). This again indicates a poor equalization, but the agreement in wave forms is better in the arrangement given in the upper part of Figure 5. This means that the north-

TABLE 1. Fault-Plane Solutions of Kern County Aftershocks [*Bâth and Richter, 1958*]

No.	Date	Origin time, GCT	deg (W)	Epicenter		deg (N)	min	<i>M</i>	Direction of strike, deg	Type of fault
				min						
	1952									
75	July 23	0039	118	35		35	22	6.1	N 50 E	Dip slip
117	July 25	1910	118	30		35	19	5.7	N 45 E	Left-hand strike slip
118	July 25	1943	118	30		35	19	5.7	N 31 E	Left-hand strike slip
141	July 29	0704	118	51		35	23	6.1	N 53 E	Right-hand strike slip
194	Aug. 22	2241	118	55		35	20	5.8	N 49 E	Right-hand strike slip

ward motion at Weston corresponds to the westward motion at Resolute Bay, again in agreement with the result obtained before.

Rayleigh waves. In his stimulating proposal, *Tukey* [1959] showed a cross-correlation function between Rayleigh waves from two different shocks, for which the wave path, epicentral distance, and recording instrument are common. This function tells us whether the senses of motions at the two sources are the same or opposite. If we know, however, the theoretical impulse response for the wave media including the instrument, we can determine the sense of motion at the source not relatively, but absolutely. As shown in a previous section, the cross-correlation function between the actual record and the theoretical impulse response gives the source function and tells us the sense of motion at the source; for example, upward or downward.

Impulse response for Rayleigh wave media. The phase-velocity curves of Rayleigh waves were obtained by *Ewing and Press* [1959] for each region of the United States from the records of a Samoa earthquake by the use of tripartite nets of stations covering the whole United States. Therefore, we can get the impulse response for Rayleigh waves for wave paths within the United States.

In this paper, Palisades and Weston records of Rayleigh waves from five Kern County aftershocks of 1952 (1 dip slip, 2 left-hand strike slips, 2 right-hand strike slips) were used. Table 1 shows their epicenters and origin times obtained by *Richter* [1955] and the fault-plane

solutions by *Bâth and Richter* [1958]. The instruments at both stations are Benioff long-period seismographs, sensitive to vertical motion. The period of the pendulum is 1 sec and that of the galvanometer is 75 sec at both stations. The damping is critical at Weston and $\epsilon_s = 1.5 \omega_s$, $\epsilon_g = 6 \omega_g$ at Palisades.

The impulse responses for Rayleigh waves at both stations from the epicenter at Kern County were computed. The total wave path was divided into seven portions (Fig. 6), and the phase delay time for each region was obtained by the use of *Press's* phase-velocity curve with the parameter *H* (crustal thickness) given by *Ewing and Press* [1959]. The total phase delay time was obtained by summing up the phase delay time for each region and adding the instrumental phase delay time. In computing the impulse response according to equation (6), ω_1 was taken as 0.016 cps and ω_2 0.07 cps, and the frequency range between them was divided into nine portions (equation 7).

The resultant theoretical seismograms for both stations are shown by dotted lines in Figure 7. Although the difference in the epicentral distance between the two stations is small (about 200 km), there is an appreciable difference between them in the wave forms at the shorter periods. In the same figure, actual seismograms smoothed by a symmetric moving average are shown in the opposite sense to the theoretical. These seismograms are from the earthquake 75, which is a dip slip. The original seismograms are shown in Figure 8.

The agreement between the actual and the

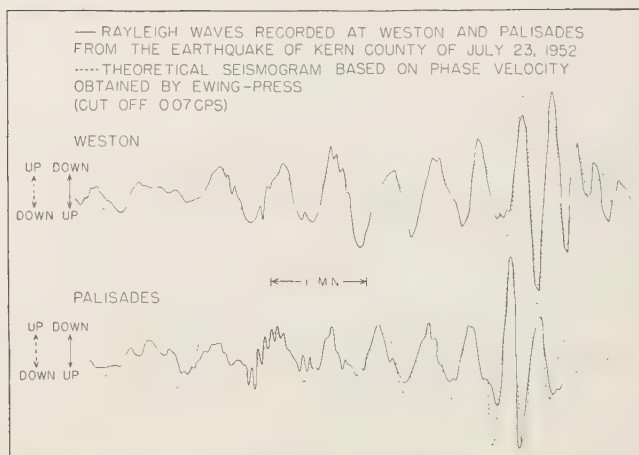


Fig. 7. Theoretical seismograms of Rayleigh waves for the wave paths from Kern County to Weston and Palisades, compared with the actual seismograms.

theoretical records is excellent at both stations. The theoretical amplitude variation follows the actual one very closely, even though we did not take the dissipation factor into account.

Source functions from Rayleigh waves. The source functions are computed from the actual records and the theoretical impulse response by the use of equation (4). Figure 9 shows the source function for the five shocks listed in

Table 1 obtained from the Weston records. Figure 10 shows those obtained from the Palisades records.

The source function for shock 75 is shown at the top in Figs. 9 and 10. This shock was initially a dip-slip motion, according to Bailey and Richter. The fault line runs in the direction $N 50^\circ E$, and the southeast side of the fault moved downward relative to the north

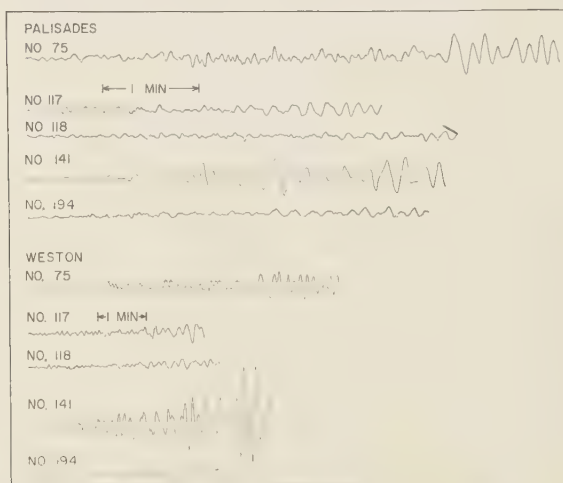


Fig. 8. Original seismograms of Rayleigh waves from Kern County aftershocks recorded at Weston and Palisades. The number is the serial shock number assigned by Richter [1955].

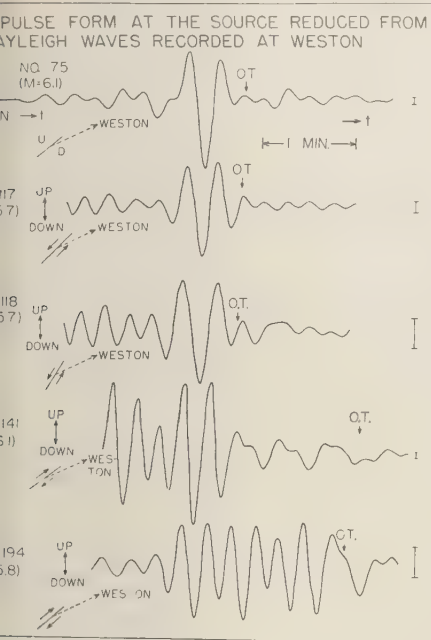


Fig. 9. Source functions for Kern County aftershocks derived from the Weston records of Rayleigh waves.

side. Both Weston and Palisades lie on the southeast side of the fault. The source function has a nearly symmetrical form with respect to the time of maximum amplitude which is downward, in agreement with the fault motion obtained from the P-wave data. A weakly oscillatory form of the source function merely means that the spectrum of recorded waves is wide enough to recover the sharper impulse function.

The time mark *O.T.* in Figures 9 and 10 is the origin time obtained from equalization. This time coincides with the time of the maximum amplitude shifted by about 20 sec. This means that the error in group velocity used in the computation of the impulse response is about 2 percent, which is not unexpected. This error does not affect the wave form very much because the wave form depends more strongly on the difference in phase delay time among frequencies.

The next two source functions in Figures 9

and 10 are for the shocks of left-hand strike slip type. They show very similar features to that of shock 75. The principal motion (the motion with the maximum amplitude) is again downward, and the time of this motion relative to the origin time is the same as that for shock 75. The fault-plane solution from the P-wave data shows that the sense of motion at the source in the directions to Weston and Palisades is toward the station for both earthquakes. According to *Lamb* [1904], a horizontal impulse applied at a point on the free surface in the sense toward the detector will give a vertical-component motion that is principally downward. Therefore, again the sense of motion at the source derived from Rayleigh waves is in agreement with the fault-plane solution from the P-wave data.

It is interesting to compare the amplitudes of the source function derived from Rayleigh waves with the amplitudes of Love waves for different types of the fault motion. Love waves from the three shocks described above are

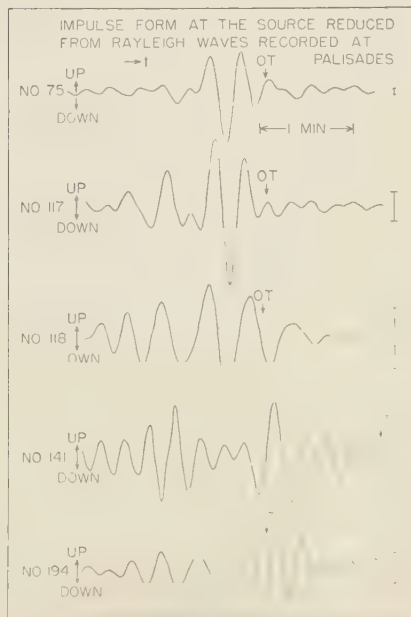


Fig. 10. Source functions for Kern County aftershocks derived from the Palisades records of Rayleigh waves.

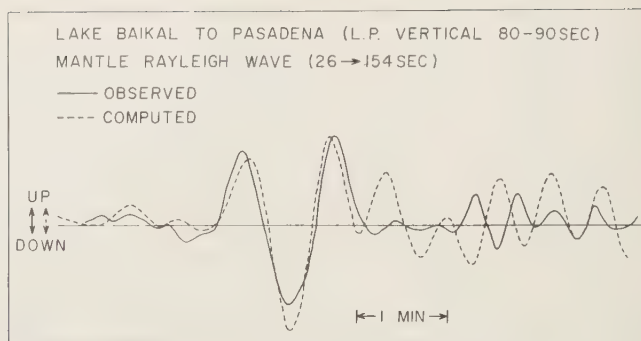


Fig. 11. Mantle Rayleigh waves from the Lake Baikal earthquake of August 29, 1959, recorded at Pasadena and the corresponding theoretical seismogram.

beautifully recorded at Weston, and their amplitudes are almost the same [Aki, 1960]. On the other hand, the source amplitude of shock 117 is 3/10 that of shock 75 from Palisades Rayleigh waves and 4/10 from Weston Rayleigh waves. The amplitude of shock 118 is even less, 1/10 that of shock 75 from Palisades Rayleigh waves and 2/10 from Weston Rayleigh waves. (The source functions in Figs. 9 and 10 are drawn in arbitrary scale in amplitude, and the length of the solid line on the right of each source function indicates a unit amplitude which is common to all the functions for the same station.)

A disturbing motion appears in the source function obtained from the Palisades record of shock 117. Since this occurs before the principal motion, it cannot be due to the event at the source. It may be due to Rayleigh waves of a secondary nature which are converted from other kinds of waves by the presence of obstacles in the wave path. This disturbance is more severe in the source function for the shocks of right-hand strike slip type which occurred in the northwest side of the White Wolf fault of the Kern County main shock.

In the case of shock 194, we cannot tell the sense of principal motion from the source functions derived either from the Palisades record or from the Weston record.

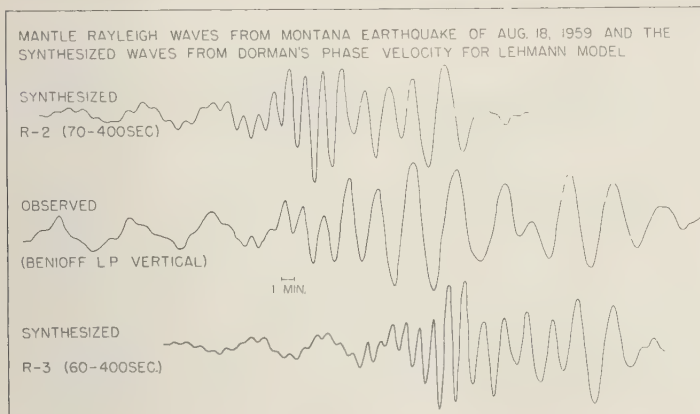
In the case of shock 141, the source function from Palisades shows a pulslike form, of which the sense of principal motion agrees with the fault-plane solution from the P-wave data. However, the time of this principal motion ap-

pears a little too early relative to the origin time. This may be due to some disturbance and the pulslike form may have been shaped accidentally.

The success in forming a clear impulse in the source function seems to depend on the position of Rayleigh waves relative to the disturbance noises in the period range from 20 sec to about 40 sec. It is not safe to use periods shorter than 15 sec because they are subject to scattering. The second mode becomes prevailing for the longer periods. Also, in order to get a well-shaped pulse form, it is necessary to have a spectral width of at least one octave.

Mantle Rayleigh waves. It is well known that the Rayleigh waves in the period range from 40 sec to 150 sec take a pulslike form due to the group-velocity maximum at a period of about 75 sec [Ewing, Jardetzky, and Press, 1957]. Figure 11 (solid curve) shows a good example of this wave. This came from the Lake Baikal earthquake of August 29, 1959, and was recorded by a long-period vertical seismograph at Pasadena. The great-circle path from Pasadena was mostly within the continental area.

A theoretical impulse response corresponding to the path discussed above is computed from the revised phase-velocity curve by Press [1960] for periods shorter than 70 sec and from one computed by Dorman and others (in preparation) based upon Lehmann's data on S-wave velocities for the longer periods. The instrumental delay is taken into account and the periods shorter than 26 sec as well as those longer than 154



12. Mantle Rayleigh waves from the Montana earthquake of August 18, 1959, recorded at Pasadena and theoretical seismograms for R2 and R3.

cut off. The resultant theoretical seismograms are shown by a dotted curve in Figure

This theoretical seismogram was computed from an upward impulse at the source. In spite of that, the pulslike form of the theoretical seismogram shows a downward principal motion. Since the actual record agrees very well with the theoretical, except for short periods, the arrangement in Figure 12 in which both have the same sense, the fault motion of this shock on the side of Pasadena must be upward if the motion is vertical and it must be away from Pasadena if the motion is horizontal.

Because of the simplicity of their wave motions, the use of mantle Rayleigh waves in the period range of 26 to 154 sec seems to be most promising for the study of earthquake mechanism. A diagnostic look at the record and a comparison of it with a set of theoretical seismograms for different epicentral distances will enable us to find out the sense of the motion at the source without a separate computation for each earthquake. Another advantage of mantle Rayleigh waves is that the mantle is more uniform than the crust and a smaller number of phase-velocity curves will cover the whole spectrum.

The mantle Rayleigh waves of even longer periods, up to 400 sec, might be useful for the study of earthquake mechanism. A tentative comparison of the actual record of the Mon-

tana earthquake of August 18, 1959, with the theoretical impulse responses for R2 and R3 is shown in Figure 12. The theoretical seismograms are obtained by the use of the phase-velocity curve by *Dorman and others* (in press) based upon Lehmann's S-wave velocity data. The actual seismogram is recorded by *Benioff's* [1959] long-period vertical seismograph with a passive network.

Although the theoretical and the experimental seismograms agree in their general features, it is difficult to find out the sense of motion at the source for both R2 and R3.

Since R2 and R3 will give the source functions at both sides of the fault, further studies of these waves will be important for the understanding of the mechanism of an earthquake.

Acknowledgments. Dr. Frank Press suggested that the writer use surface waves for the study of earthquake mechanism. The writer is most grateful for his original suggestion as well as for his valuable advice during the course of the present study. The writer's thanks are also due Drs. Hugo Benioff and Charles F. Richter for their helpful advice and suggestions. His thanks also go to the seismologists at the Resolute Bay, Weston, and Palisades stations who furnished the data on which this study is based, and he acknowledges the help of John M. Nordquist in preparing the program for the Bendix G-15 computer at our laboratory.

This work was partially supported by grants from the Interdisciplinary Research Program of

the International Geophysical Year and the Office of Ordnance Research.

REFERENCES

- Aki, K., The use of Love waves for the study of earthquake mechanism, *J. Geophys. Research*, 65, 323-331, 1960.
- Båth, M., and C. F. Richter, Mechanism of the aftershocks of the Kern County, California earthquake of 1952, *Bull. Seis. Soc. Amer.*, 48, 133-146, 1958.
- Benioff, H., Passive networks for seismographs, Reported at the annual meeting of the Seismological Society of America, April, 1959.
- Brune, J., J. Nafe, and J. Oliver, A simplified method for the analysis and synthesis of dispersed wave trains, *J. Geophys. Research*, 65, 287-304, 1960.
- Dorman, J., Numerical solutions for Love wave dispersion on a half-space with double surface layer, *Geophysics*, 24, 12-29, 1959.
- Dorman, J., W. M. Ewing, and J. Oliver, Study of shear velocity distribution in the upper mantle by mantle Rayleigh waves, in press.
- Ewing, W. M., and F. Press, Determination of crustal structure from phase velocity of Rayleigh waves; Part III: The United States, *Bull. Geol. Soc. Am.*, 70, 229-244, 1959.
- Ewing, W. M., W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media*, McGraw-Hill, New York, 357 pp., 1957.
- Knopoff, L., and F. Gilbert, First motions from seismic sources, *Seismic Scattering Project*, Fourth annual report, sect. II, chap. X, Inst. Geophysics, Univ. of California, 1959.
- Lamb, H., On the propagation of tremors over the surface of an elastic solid, *Phil. Trans. R. Soc., London, Ser. A*, 203, 11, 1904.
- Press, F., Crustal structure in California-Nevada region, *J. Geophys. Research*, 65 (in press), 1960.
- Richter, C. F., Foreshocks and aftershocks; Earthquakes in Kern County during 1952, *Calif. Dep. Nat. Resources, Div. Mines, Bull.* 171, 177-195, 1955.
- Sato, Y., Analysis of dispersed surface waves; II, and III, *Bull. Earthquake Research Inst. Tokyo Univ.*, 33, 33-48, 1955; 34, 9-18, 1956.
- Tukey, J., Equalization and pulse shaping techniques applied to the determination of initial sense of Rayleigh waves, Report on a Panel Seismic Improvement, Appendix 9, (Chairman L. V. Berkner), 1959.
- Valle, P. E., Sulla misura della velocità di gruppi delle onde sismiche superficiali, *Ann. geofis. Rome*, 2, no. 3, 1949.

(Manuscript received October 19, 1959.)

Apparatus for Phase-Equilibrium Measurements at Pressures up to 50 Kilobars and Temperatures up to 1750°C

F. R. BOYD AND J. L. ENGLAND

*Geophysical Laboratory
Carnegie Institution of Washington
Washington, D. C.*

Abstract. Construction and calibration of apparatus utilizing a solid pressure medium for phase-equilibrium studies at elevated temperatures and pressures are described. Pressure calibration is carried out by measurement of the Bi I Bi II and Tl II Tl III transitions. A new determination of the thallium transition, 37.1 ± 1.3 kilobars, is given. Tests indicate that talc is superior to pyrophyllite and boron nitride as a solid pressure medium for high-temperature work.

INTRODUCTION

The apparatus described below was developed to permit study of phase equilibria in the ranges of pressure and temperature present in the upper part of the earth's mantle. Such studies can help define and solve a variety of geophysical and petrological problems. The chemical and mineralogical constitution of the mantle and the nature of the Mohorovicic discontinuity are matters of great current interest. Many lavas have originated in the mantle, and their temperatures of formation and compositions must have been influenced by high pressure. Phase-equilibrium determinations at high temperatures and pressures can put limits on the mineralogical nature of the mantle and may provide data on the melting relations at these pressures that will be of considerable significance in understanding the origins of lavas. Although this paper covers only the construction and calibration of the apparatus, some results on the quartz-coesite transition are presented in the following paper in this issue.

Studies of silicate equilibria at high temperatures and at pressures in excess of 10 kb are most conveniently carried out with a pressure system utilizing a solid pressure medium. Coes [1955] developed a simple, internally heated high-pressure system of this sort for use up to 2000°C and 45 kb. (The details of his apparatus have not yet been published.) Hall [1958] has shown that this design may be modified to reach temperatures above 2000°C. The appa-

ratus described in this paper is based on the Coes-Hall design. Developments made by the present authors include chiefly the method of introducing thermocouple leads into the pressure chamber. This apparatus can be used up to a pressure of about 50 kb at a temperature of 1750°C. The measurable temperature range may be extended by using thermocouples that melt at higher temperatures than platinum/platinum-10 per cent rhodium.

APPARATUS

The apparatus consists of an internally heated tungsten carbide pressure vessel supported by a steel ring. The pressure vessel is end-loaded during a run with a thrust of 150 to 250 tons delivered by a hydraulic press. Pressure is applied to the run and furnace assembly with a carbide piston driven by a second hydraulic press. The assembly is illustrated in Figure 1.

The carbide core of the pressure vessel is ground in the form of a tapered cylinder approximately 2 in. long, 2 in. in diameter, and with a bore of 0.500 in. The taper on the outside of the carbide cylinder is 1° included angle. The core is forced into a steel supporting ring with an interference of 0.018 to 0.022 in. on the diameter. The steel supporting ring has an outside diameter of 6¼ in. and is jacketed with a soft-steel safety ring ¾ in. thick.

We have used Carboloy grade 883 for the core of the pressure vessel. A variety of steels have been employed for the supporting ring;

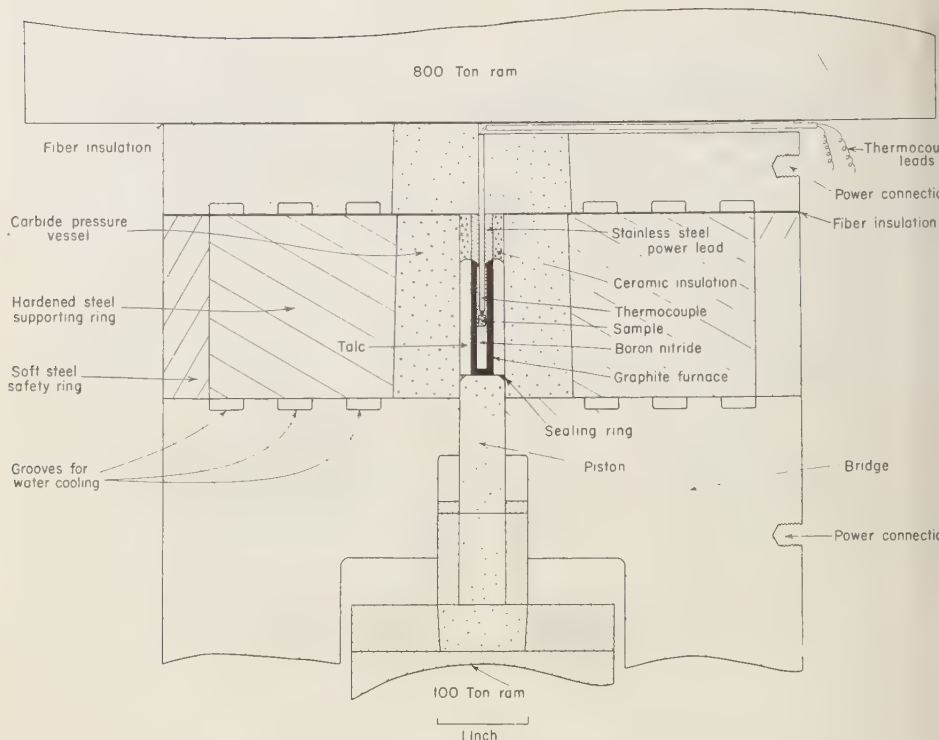


Fig. 1. Apparatus for use in the pressure range up to 50 kb at temperatures up to 1750°C. Carbide parts are stippled; steel parts are ruled.

satisfactory results have been obtained using AISI E4340 forged as rings and hardened to Rockwell C44-46. The bores of the 4340 rings are stretched about 1 per cent on a steel mandrel after hardening. Autofrettaging a ring substantially increases its yield point.

The core of a pressure vessel constructed in this manner ordinarily lasts up to 60 runs at high temperature. Fracture starts in the first few runs with the development of radial cracks in the central part of the bore and with cracks normal to the axis of the bore forming opposite the ends of the piston and base plug. These fractures fill up with lead (used as a jacket on the furnace assembly, Fig. 2) during a run and do not interfere with the operation of the apparatus. Eventually chips spall off the end of the core adjacent to the base plug, and the end loading becomes less effective. When the ends of a core are well chipped the lateral cracks in

the bore can expand and the core becomes unusable. The core can then be pressed out of the ring fitted with a new one.

Pistons can be made conveniently from standard carbide inserts 0.500 in. in diameter by 1 in. long. The bore of the pressure vessel ground to a diameter 0.0005 to 0.001 in. larger than the piston. The end of the piston that bears on the ram of the press is jacketed with steel to prevent chipping. Pistons last indefinitely at pressures below 30 kb; they last up to 40 runs in the range 30 to 50 kb.

Grooves are cut in the base plate and bridge adjacent to the faces of the pressure vessel, permit water cooling. Without water cooling the steel supporting ring expands away from the carbide core on high-temperature runs and causes its rapid deterioration.

The furnace and base plugs assembly is illustrated in detail in Figure 2. The run in

FURNACE ASSEMBLY

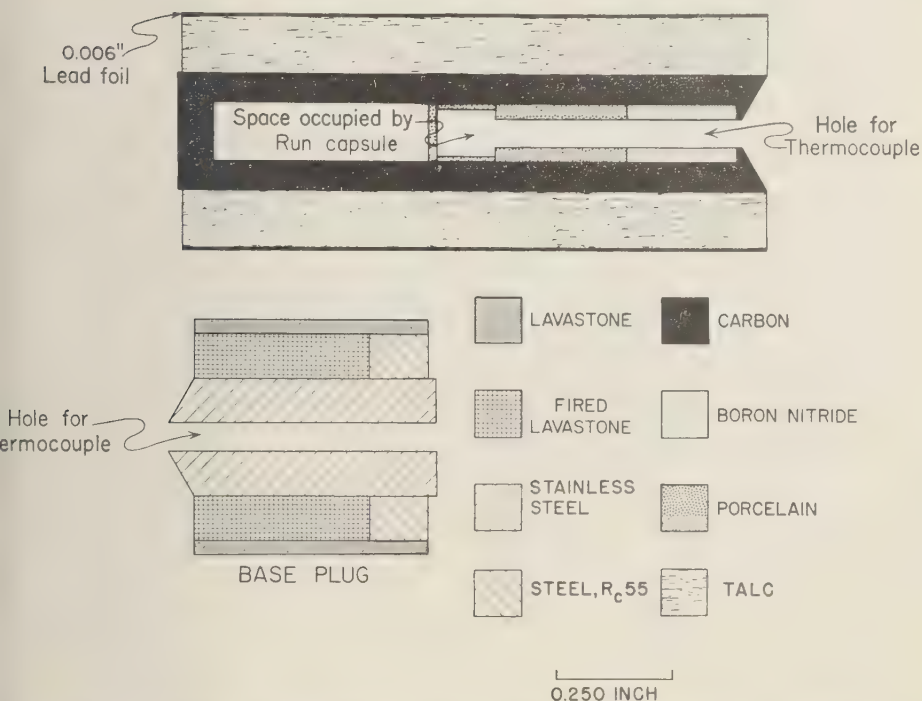


Fig. 2. Detail of furnace and base plug assembly used in apparatus shown in Figure 1. The assembly is not recoverable, and a new unit is used for each run.

of a powder is loaded in a platinum capsule with a wall thickness of 0.003 in. The run capsule are then compressed in a pellet press to a diameter of 0.096 in. and a length of 0.090 to 0.125 in. The furnace is a graphite tube, 1/4 in. in outside diameter, with a bore of 0.096 in. and a length of 1 1/8 in. Talc is used for insulating sleeve and pressure medium. Boron nitride and high-temperature porcelain are used for inserts within the furnace. Boron nitride, which is isostructural with graphite, is found to be soluble in platinum and lowers its melting point by an amount on the order of 100°C. Hence the thermocouple and platinum capsule around the run are shielded from contact with boron nitride by porcelain inserts.

The base plug consists of a stainless-steel washer lead insulated by a ceramic sleeve. The power lead can be made conveniently from pressure tubing, but it must be annealed be-

fore using. The thermocouple is introduced in a ceramic tube through a hole in the power lead and is held in place by friction. The hardened-steel washer at the base of the assembly is designed to reduce stress at the vulnerable edge of the bore. Fired pyrophyllite tends to chip the bore on release of pressure; hence the base-plug assembly is surrounded with a thin sleeve of unfired pyrophyllite. For consistent results clearances between the various parts of the furnace and base-plug assembly must be held to about 0.001 in.

The power supply consists of a 220-volt 50-amp variable transformer, the output of which is fed through a 5-kw transformer with a winding ratio of 20:1. Maximum power output is thus about 500 amp at 10 volts. Fine adjustment of the power is achieved by adding a small voltage, in phase, through a second variac. The furnace illustrated in Figure 2 has very

little thermal inertia, but it has proved easy to control the temperature manually to $\pm 5^\circ$, even at 1700°C . The power required to maintain a temperature of 1700°C is about 1.8 kw. The small thermal inertia of the furnace contributes to a fast quench. When the power is shut off, the temperature of the run drops to the temperature of the pressure vessel wall ($< 500^\circ\text{C}$) within 5 seconds.

PRESSURE CALIBRATION

The pressure on a run in this apparatus is determined by measuring the oil pressure in the hydraulic press and computing a load pressure from the known cross-sectional areas of the piston and hydraulic ram. The computed pressure is then corrected for friction in the press and high-pressure assembly. A 1400-bar Heise gage with an accuracy of better than 0.1 per cent is used to measure the oil pressure. The friction is found by calibration at known transition points. The largest part of the friction is due to the shear strength of the pressure medium; mechanical friction in the hydraulic press and the pressure assembly is very small.

The problem of evaluating the friction is made difficult by the fact that there is no calibration point in the pressure range 10 to 50 kb at high temperature. Since, however, the shear strengths of solid pressure media decrease with temperature, values of friction measured at room temperature form upper limits for the friction actually present at high temperatures. The friction in the hydraulic press and piston assembly itself is independent of run temperature and may be taken as a lower limit. As is described below, it is possible to make the difference between these limits as small as 10 per cent.

There are two calibration points at room temperature in the pressure range 10 to 50 kb that can conveniently be used. These are the transition Bi I-Bi II at about 25 kb and the transition Tl II-Tl III at about 37 kb.

The transition Bi I-Bi II has been located by Bridgman [1940] at 24.9 kb at 30°C and used by him as a primary calibration point. There is a second transition in bismuth (Bi II-Bi III) at 26.4 kb that is less convenient to measure. Bridgman [1935] located a transition in thallium by volume discontinuity at 41,000 kg/cm²

(40.2 kb) at 30°C . The phase diagram presented by Bridgman [1935, p. 899] identifies this transition as Tl II-Tl III. In a later study with an anvil apparatus Bridgman [1936, p. 208] found this transition occurring with increasing pressure at 45,000 kg/cm² (44.1 kb), but [1936, p. 208] declined to choose between the values.

In the course of a study of solid pressure media a silver chloride cell was developed which yielded a remarkably low hysteresis for the I-Bi II transition. The agreement with Bridgman's bismuth point is excellent, and this method has been used to obtain a more accurate value for the Tl II-Tl III transition.

Bismuth transition. Two samples of bismuth were used. One of them, obtained from the National Bureau of Standards through Alvin Valkenburg, is 99.99 per cent pure. The other is the electrolytic bismuth used by Bridgman. These samples gave identical results. The termination was made with the same piston and pressure vessel assembly described above, but with a silver chloride cell $\frac{1}{8}$ in. thick substituted for the talc furnace assembly (Fig. 1). Embedded in the silver chloride was a bismuth wire 0.013 in. in diameter by $\frac{3}{8}$ in. long. The wire was connected by gold leads to the piston and to an insulated base plug. The transition was detected by the change in electrical resistance.

A run across the transitions Bi I-Bi II and Bi II-Bi III is illustrated in Figure 3. The hysteresis shown

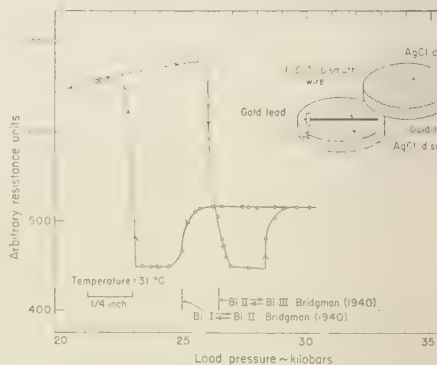


Fig. 3. The transitions Bi I \rightleftharpoons Bi II \rightleftharpoons Bi III as shown by the change in electrical resistance of bismuth wire in an AgCl cell.

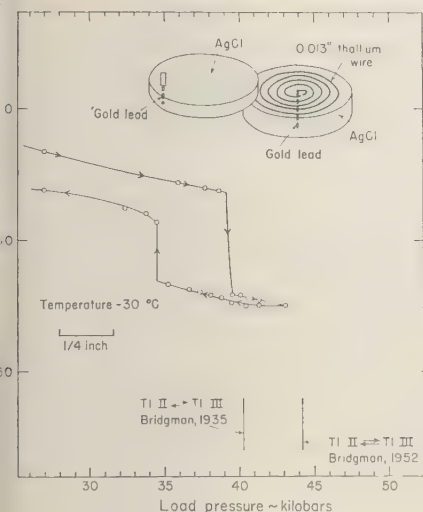


Fig. 4. The transition $Tl II \rightleftharpoons Tl III$ as shown by the change in electrical resistance of a thallium wire in an $AgCl$ cell.

ure 3 for the transition I-II is 11.6 per cent. By balancing on the transition, with I and II present, increasing the pressure until $I \rightarrow II$ and then releasing the pressure until $II \rightarrow I$, it was possible to reduce the hysteresis to 3.1 per cent at a mean value of 25.2 kb, in good agreement with Bridgman's value of 24.9 kb.

Thallium transition. The II-III transition in thallium is sharp, and it occurs at a pressure sufficiently higher than the bismuth transitions to make it of interest as a calibration point. The chief difficulty in working with the thallium point is that the change in electrical resistance is only 28 per cent [Bridgman, 1952, p. 100], smaller by a factor of 20 than the change in the Bi I-Bi II point.

One sample of thallium was obtained, through the gift of Van Valkenburg, from the National Bureau of Standards. It is 99.9 per cent pure. A second sample was obtained from the Fisher Scientific Company. It is listed as 'purified.' Savings of these materials were extruded through a die into Nujol to form 0.013-in.-diameter wires. It was necessary to use $2\frac{1}{2}$ in. of wire to obtain a sufficiently large resistance change. For establishing the calibration point a silver chloride cell sketched in Figure 4 was used. The thallium wire was wound in a spiral

TABLE 1. The Transition $Tl II \rightleftharpoons Tl III$
Temperature $29^\circ \pm 1^\circ C$

Thallium	Pressure,* kb	Hysteresis, %
NBS	37.15	6.8
NBS	37.10	7.2
Fisher	37.15	8.7

* Mean of the pressures at which the transition started on increasing and decreasing the load.

between two disks of silver chloride. The change in resistance with pressure on one run is shown in Figure 4. By balancing on the transition in the manner described above for bismuth it was possible to reduce the hysteresis below that shown.

Results obtained for three set-ups, two with the National Bureau of Standards thallium and one with the Fisher thallium, are given in Table 1. The agreement of the means on the three set-ups is excellent, and there is clearly no measurable difference between the two samples of thallium. From these data the $Tl II$ - $Tl III$ point can be taken to be $37.1 \text{ kb} \pm 3.5$ per cent.

PRESSURE MEDIA

A pressure medium for high-temperature runs not only must have a low shear strength but also must be an electrical insulator and a good thermal insulator. Silver chloride is by far the best solid pressure medium tried by the authors at room temperature, but it melts at $455^\circ C$ (at 1 atm) and becomes an electrical conductor. At high temperatures other materials must be used.

To explore the suitability of various solid pressure media we have used cells of the sort illustrated in Figure 5. These cells, patterned after a design by Hall [1958, Fig. 3], give values of friction equivalent to those obtained with actual furnace assemblies. Data are presented below for talc, pyrophyllite, and boron nitride. The talc and the pyrophyllite, obtained from the American Lava Corporation in the form of pressed cakes, are respectively Lava Grade 1136 and Lava Grade A. The boron nitride was obtained in similar form from Union Carbide Corporation.

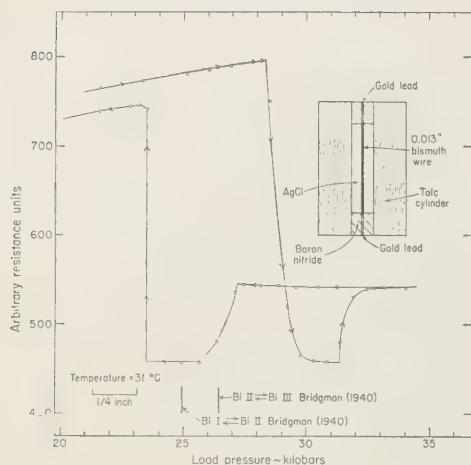


Fig. 5. The transitions $\text{Bi I} \rightleftharpoons \text{Bi II} \rightleftharpoons \text{Bi III}$ as shown by the change in electrical resistance of a bismuth wire in a talc cell.

A typical run with bismuth in a talc cell is illustrated in Figure 5. The point at which Bi I starts to invert to Bi II is sharp and reproducible to about ± 0.5 per cent of the pressure. Measurement of this point is unaffected by holding the cell at a pressure slightly below the transition pressure for periods up to half an hour. It can be noted in Figure 5 that the hysteresis loop is not symmetrical about the true inversion point. This phenomenon is characteristic of these pressure media; it results from the fact that a part of the deformation is elastic rather than plastic. The phenomenon may be understood by imagining the sleeve of pressure medium as a spring which must be compressed in bringing pressure to bear on the bismuth wire; when the pressure is released the 'spring' snaps back, causing an unsymmetrical hysteresis loop. With some set-ups we have observed the reverse reaction $\text{Bi II} \rightarrow \text{Bi I}$ proceeding at a load pressure actually slightly higher than the true transition pressure. With these media the pressure at which $\text{Bi II} \rightarrow \text{Bi I}$ is not so reproducible as the advance reaction and is dependent on how long the sample is held above the transition pressure.

Data on identical cells of talc, pyrophyllite, and boron nitride are presented in Table 2. These cells were jacketed with 0.006-in.-thick lead foil. The listed frictions are the difference

TABLE 2. Values of Friction Obtained with Various Pressure Media at the Transition $\text{Bi I} \rightarrow \text{Bi II}$ 24.9 kb, temperature 30°C

Pressure Medium	Friction on Advance
Talc A	13
B	13
C	14
Boron nitride	20
Pyrophyllite	26

between the true transition pressure (24.9 kb) and the load pressure at which $\text{Bi I} \rightarrow \text{Bi II}$ expressed as a percentage of the true transition pressure. The blocks of talc from which the cells were cut show a slight grain. The talc cells (A, B, and C in Table 2) were from mutually perpendicular slices in a block to test any possible effect of grain; no significant effect was found. These data show that talc is considerably better than boron nitride or pyrophyllite as a pressure medium. Talc has an additional advantage over pyrophyllite in that it is stable to higher temperatures.

A talc cell has also been tested at the thallium transition. The friction at the point $\text{Th I} \rightarrow \text{Th II}$ measured relative to the equilibrium point (37.1 kb, see above) is 13.6 per cent, in excellent agreement with the results obtained with bismuth. It is evident that at least over the range 20 to 40 kb the friction in a talc cell is a constant percentage of the load pressure.

To be certain that the type of cell illustrated in Figure 5 adequately models a high-temperature furnace assembly a run was made with an actual furnace assembly in which the sample and platinum capsule were replaced with a cylinder of AgCl the same size as the run containing a $\frac{1}{8}$ -in. length of bismuth wire. No other changes were made. The friction at the $\text{Bi I} \rightarrow \text{Bi II}$ point was found to be 13.4 per cent, in agreement with the results in Table 2.

For runs at high temperatures a correction of -13 per cent to the load pressure is taken as one limit and a correction of -3 per cent (estimated minimum friction) as the other. These limits involve two assumptions: that the friction should be counted as negative, a

at high temperature the friction decreases from its room-temperature value.

During high-temperature runs in this apparatus the piston continually advances, and pumping is necessary to maintain the pressure. Results from creep in the base plug assembly and from extrusion of lead into cracks in the wall of the pressure vessel. At pressures above 15 kb these effects dominate the temporary effect of thermal expansion as the furnace is heated to temperature. Since the piston is continually advancing on the furnace assembly, a correction for friction should be negative.

At high temperature the shear strengths of these pressure media decrease, and, hence, so should the friction. This effect may be slightly offset by chemical changes. At temperatures above 1000°C a thin zone of the talc adjacent to the graphite furnace breaks down to talc and quartz (or coesite). At a run temperature of 1700°C this zone has a maximum thickness of 0.018 in. It is unlikely that this shell could be strong enough to materially influence the pressure transmitted to the run. Moreover, at high temperature the enstatite and quartz are crystallizing in a hydrous environment (H_2O given off by the talc) and may have essentially no strength.

TEMPERATURE MEASUREMENT

Temperature measurement in high-pressure apparatus is subject to error due both to gradients within the apparatus and to the effect of pressure on the temperature-sensing element. There are no direct measurements of the effect of pressure on the emf of a thermocouple in the ranges of temperature and pressure utilized with this apparatus. Measurements at lower temperatures and pressures and some indirect experiments, reviewed below, indicate that the effect of pressure on the emf of a platinum/platinum-10 per cent rhodium thermocouple is small, and the effect is probably negligible in comparison with the precision of measurement now attainable. Error due to thermal gradients can be more directly evaluated.

We have measured the temperature gradient in a furnace of the design in Figure 2 in a series of runs by introducing two platinum/platinum-10 per cent rhodium thermocouples with their junctions either 1/8 or 1/16 in. apart

and arranged as nearly as possible along the axis of the furnace. In these experiments one of the junctions was placed in the same position in the furnace as the thermocouple on a routine run; the position of the other, depending on the set-up, corresponded to either the center or the opposite end of a run capsule. The differences between the couples, therefore, approximate the gradient across an actual run.

As can be seen in Figure 6 the maximum difference found was 15°. Making no correction to the reference couple reading, runs can be assigned an uncertainty of $\pm 10^\circ$ below 1500°C and $\pm 15^\circ$ above.

Birch [1939] has examined the effect of pressure on chromel-alumel and platinum/platinum-10 per cent rhodium thermocouples at pressures up to 4000 bars and temperatures up to 500°C. Birch found no effect for the chromel-alumel couple, but measured a small decrease in the emf of the platinum couple with increasing pressure. The correction for the platinum couple was found to be a linear function of both temperature and pressure, reaching a maximum value in his measurements of $15\mu\text{V}$ (1.5°) at 470°C and 4100 bars. Extrapolation of Birch's data to higher temperatures and pressures can be expected to give a measure of the maximum possible correction, since most pressure effects tend to decrease with increasing

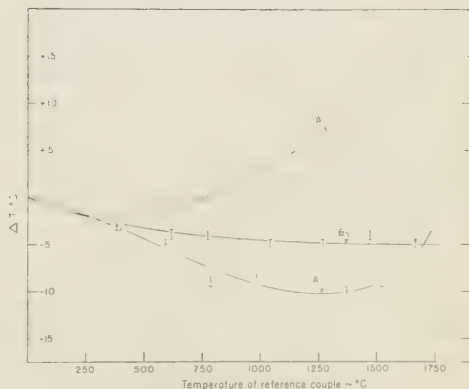


Fig. 6. Temperature gradients measured in the furnace assembly shown in Figure 2. For runs *A* and *A'* the reading couple was mounted $\frac{1}{8}$ in. from the reference couple; for run *B* the reading couple was $\frac{1}{16}$ in. from the reference couple.

pressure. The correction at 40,000 bars and 1700°C would be +48°.

Hall [1955, p. 1145] compared a platinum/platinum-10 per cent rhodium couple and a chromel-alumel couple at temperatures up to 900°C and pressures up to 100,000 atm. He found an average deviation from the mean temperature of $\pm 3^\circ$ at 900°C. We have repeated Hall's experiment up to a pressure of 45 kb. A measurable effect was found; its magnitude is in rough agreement with Hall's data and is substantially less than that predicted by extrapolation of Birch's measurements.

A chromel-alumel couple and a platinum/platinum-10 per cent rhodium couple were run with their junctions, separated by ceramic approximately 0.030 in. thick, placed in the position normally occupied by a run capsule. Changes in emf of the two couples were then read as a function of temperature and pressure. Two set-ups were made with a number of runs. At high temperature and pressure the chromel-alumel couple read a higher apparent temperature than the platinum couple. The maximum difference was $10^\circ \pm 1^\circ$ at 955°C and 45.5 kb. An average of three runs at about 950°C gave a change with pressure in the range 17 to 45.5 kb of $0.27^\circ/\text{kb}$; one run at 500°C in the same pressure range gave $0.07^\circ/\text{kb}$. The two couples, therefore, read measurably different at high temperature and pressure, and the difference increases with temperature and pressure.

The experiment described above indicates a small difference between the two types of thermocouple but yields no positive information about the pressure effect on an individual couple. In making these runs, we repeatedly set the potentiometer on one of the couples and rapidly increased the pressure without adjusting the power input. The platinum/platinum-10 per cent rhodium couple showed either no change or a small erratic drift; the chromel-

alumel couple invariably drifted to a high emf. This suggests that the difference in reading between the couples is largely due to pressure effect on the chromel-alumel couple.

Clearly, present data are too limited to make a meaningful correction for the effect of pressure on the emf of a thermocouple at pressures above 10 kb and temperatures above 1000°C. Hall's data and ours, nevertheless, indicate that the correction will be of the same order as the present precision of temperature measurements in this pressure range.

Acknowledgments. The construction of much of the apparatus described above was carried out with skill and ingenuity by Mr. O. R. McCluney and his work is gratefully acknowledged. Dr. Alvin Van Valkenburg of the National Bureau of Standards supplied samples of pure bismuth and thallium, and Professor P. W. Bridgman sent us some of his stock of electrolytic bismuth. Drs. Sydney P. Clark and H. S. Yoder and Professor Francis Birch have read this manuscript and offered many helpful suggestions.

REFERENCES

- Birch, F., Thermoelectric measurement of high temperatures in pressure apparatus, *Rev. Sci. Instr.*, **10**, 137-140, 1939.
- Bridgman, P. W., Polymorphism, principally of the elements, up to 50,000 kg/cm², *Phys. Rev.*, **48**, 893-906, 1935.
- Bridgman, P. W., The measurement of hydrostatic pressure to 30,000 kg/cm², *Proc. Am. Acad. Arts Sci.*, **74**, 1-10, 1940.
- Bridgman, P. W., The resistance of 72 elements, alloys, and compounds to 100,000 kg/cm², *Proc. Am. Acad. Arts Sci.*, **81**, 165-251, 1952.
- Coes, L., High-pressure minerals, *J. Am. Ceram. Soc.*, **38**, 298, 1955.
- Hall, H. T., The melting point of germanium as a function of pressure to 180,000 atm, *J. Phys. Chem.*, **59**, 1144-1146, 1955.
- Hall, H. T., Some high-pressure, high-temperature apparatus design considerations: equipment for use at 100,000 atm and 3000°C, *Rev. Sci. Instr.*, **29**, 267-275, 1958.

(Manuscript received December 2, 1959.)

The Quartz-Coesite Transition

F. R. BOYD AND J. L. ENGLAND

*Geophysical Laboratory, Carnegie Institution of Washington
Washington, D. C.*

Abstract. The quartz-coesite transition curve has been determined over the temperature range 700° to 1700°C in the pressure range 20 to 40 kb. The equation for the curve is $P' = 19.5 + 0.0112T'$, where P' is in kilobars and T' is in degrees centigrade. The determination was made with an internally heated tungsten carbide pressure vessel utilizing talc for a pressure medium. The results of some runs on the quartz-coesite transition with an anvil apparatus are described. These data are in poor agreement with the determination cited above, and the difference is interpreted as an effect of pressure gradients in quartz-coesite runs in the anvil apparatus. Quartz could not invert to coesite in the earth at depths less than about 100 km, and it is therefore unlikely that coesite has formed in crustal rocks.

Introduction. The dense polymorph of silica now known as coesite was one of the first new phases discovered in investigations of silicate equilibria at high pressures and temperatures. Since the first synthesis by Coes [1953] there has been much interest in determining the properties and stability of coesite and in the possibility of finding a natural occurrence of the phase.

In discussing the quartz-coesite reaction, MacDonald [1956] pointed out that transitions in other minerals with open, framework structures are to be expected in the same pressure range. This prediction has been amply confirmed. Transitions or reactions to phase assemblages of smaller volume are now known to take place with quartz, albite [Kennedy, personal communication; Birch and LeComte, 1959] or calcic feldspar [Kennedy, personal communication], nepheline and leucite [Boyd and England, unpublished data], all in the pressure range 10 to 40 kb. It is now clear that phase relations in systems involving quartz, feldspars, and feldspathoids will be very different at high pressures from the relations found at atmospheric pressure and in hydrothermal studies. There is as yet insufficient data to assess the petrological and geophysical implications of the reactions now known, and many others have yet to be studied. Improvements in technique, however, make it possible to extend the range and accuracy of the data and permit its geological significance to be better understood. It is the

purpose of the present work to extend the determination of the quartz-coesite curve to higher temperatures and to provide an improved estimate of its position and slope.

Previous work. Coes [1953] found that coesite formed most easily in the temperature range 500° to 800°C at pressures in excess of 35 kb. He also determined the optical and X-ray properties and density of the phase. MacDonald [1956] made the first determination of the quartz-coesite reaction curve, in the temperature range 400° to 600°C. He derived the entropy and heat of formation of coesite from the curve and used the values obtained to place limits on the stability ranges of olivines and pyroxenes at high pressures. Griggs and Kennedy [1956] have briefly described a preliminary curve which slightly modifies MacDonald's determination and extends it to 900°C. Khitarov, Slutskiy, and Arsen'yeva [1957] have described the synthesis of coesite, but provide no further data on its stability range. The crystal structure of coesite has recently been determined by Zoltai and Buerger [1959].

Apparatus. The construction and calibration of the apparatus used in this investigation have been described in a previous paper [Boyd and England, 1960]. The pressure vessel consists of a cylinder of cemented tungsten carbide supported by steel rings. The run is heated by a graphite furnace, insulated by talc, inside the pressure vessel. Pressure is applied to the run, furnace, and thermocouple assemblies by a

tungsten carbide piston. The sample consists of about 25 mg of material pressed into a platinum capsule. The precision of most of the measurements is about ± 0.5 kb and ± 10 to 15°C . The accuracy of pressure measurement is ± 5 per cent, determined by calibration with the transitions Bi I-Bi II and Tl II-Tl III. The run temperatures are corrected for the thermal gradient between the thermocouple and the center of the run (5° to 40°), but no correction has been made for the effect of pressure on the emf of the Pt-Pt 10 per cent Rh thermocouple. This correction is inadequately known in the pressure range of this determination but is believed to be of the same order as the precision of temperature measurement. Owing to design changes in the course of the investigation some of the limits of error of temperature assigned to the runs discussed below are larger than those indicated possible with the assembly described in our previous paper.

Experimental procedure. Quartz, synthetic cristobalite mixed with 10 per cent silicic acid, and coesite were used as starting materials. The analysis of the quartz shows a residue after treatment with $\text{HF} + \text{H}_2\text{SO}_4$ and subsequent ignition of 0.03 per cent. Cristobalite was prepared by heating the quartz in a platinum crucible at 1500°C for 3 to 4 hours. The silicic acid is a Fisher certified reagent; it has a 'non-volatile with HF ' fraction of 0.07 per cent. Bulk samples of coesite were prepared from the cristobalite-10 per cent silicic acid mix at 1000 to 1100°C and 40 kb.

Coesite forms in 1 hour or less from dry quartz at sufficiently high pressure and temperatures in excess of 1200°C . Below 1200°C some H_2O must be present for the reaction to run in a reasonable length of time. In this series of experiments the platinum capsules containing the charges were not sealed by welding, and the H_2O slowly diffused out during the runs. In our first runs silicic acid containing about 20 per cent H_2O was used without admixture of quartz or cristobalite, and the products were found to contain some glass when quenched from temperatures as low as 900°C . The H_2O -silica fluids obtained at temperature in these runs crystallized partly to quartz in the coesite field during the quench. Subsequent runs were accordingly made with a mixture of

TABLE 1. X-Ray Data for Coesite Synthesized at 1400°C and 40 Kilobars

<i>d</i>	<i>I</i>	<i>d</i>	<i>I</i>
6.19	3	1.839	3
4.37	2	1.794	4
3.436	52	1.787	4
3.099	100	1.715	9
2.765	8	1.698	10
2.698	11	1.655	5
2.337	3	1.584	5
2.295	6	1.548	6
2.186	4	1.409	2
2.033	6	1.345	6
1.849	5		

90 per cent cristobalite and 10 per cent silicic acid. With this starting combination complete crystallization to coesite can be obtained under favorable *P-T* conditions, and no glass was found in runs at temperatures below 1200°C . This mix crystallizes to fine-grained quartz on the way to temperature and then gradually transforms to coesite if held within the coesite field. The inversion of coesite to quartz proceeds as readily as the reverse reaction. At temperatures below 1200°C it proved advantageous to mix 10 per cent silicic acid with the coesite for starting material.

The best conditions for forming coesite are temperatures above 900°C at pressures a few kilobars or more within its stability field. Runs under these conditions normally yield 100 per cent coesite. Below 900°C the rate of formation of coesite drops off rapidly, and in the vicinity of the phase boundary only small yields are obtained in runs up to 5 hours in length.

Coesite forms anhedral grains about 0.03 to 0.05 mm in diameter in runs in which it is the only phase present. It usually develops as euhedral crystals in runs in which isolated grains of coesite grow in a matrix of quartz. One run yielded euhedral crystals of coesite up to 1 mm in diameter mixed with finer-grained quartz.

All runs were analyzed by standard optical and X-ray methods. Because of the large difference in refractive indices between quartz and coesite (mean indices 1.550 and 1.595, respectively) smaller amounts of coesite mixed with quartz can be detected with the microscope than with an X-ray diffractometer.

TABLE 2. Results of Runs on the Quartz-Coesite Transition

Mix	Temperature,* °C	Pressure,† kb	Time, min	Results
Q	1710	37.3	20	Quartz
SC	1710	37.6	20	Quartz
Q	1710	39.1	20	Quartz + 10% coesite
Q	1710	40.3	17	Coesite + 40% quartz
Q	1510	35.7	17	Quartz
Q	1540 ± 40	37.0	25	Coesite
SC	1450	37.0	60	Coesite
Q	1340 ± 40	33.7	60	Quartz
Q	1340 ± 40	35.0	60	Quartz
Q	1340 ± 40	36.3	60	Coesite
SC	1210	32.5	40	Quartz + 50% coesite
C	1110	30.0	60	Quartz
C	1110	31.7	60	Quartz + 20% coesite
C	1130 ± 30	33.0	60	Coesite + trace quartz
C	1110	34.2	60	Coesite
C	1005	30.5	75	Coesite + 5% quartz
C	905	28.4	100	Quartz
SC'	905	29.7	140	Coesite + 20% quartz
C	905	31.0	90	Coesite + 50% quartz
SC'	850	30.0	130	Coesite
C'	805	28.4	200	Quartz
SC'	705	26.1	515	Quartz + 10% coesite
C	705	27.4	300	Quartz
C	705	29.4	300	Quartz + trace coesite

* Unless otherwise stated, temperatures are $\pm 10^\circ$ below 1200° and $\pm 15^\circ$ above 1200°C .

† Precision of pressure measurement and control is about ± 0.5 kb.

Q, quartz; SC, coesite; C, 90% cristobalite + 10% silicic acid; SC', coesite + 10% silicic acid; C', 80% cristobalite + 20% silicic acid.

Trace = $< 5\%$.

The d values and relative peak intensities determined for coesite synthesized at 1400°C and 0 kb are given in Table 1. These measurements were made with a Norelco diffractometer using quartz as an internal standard. Optical properties determined for a particularly coarse-grained sample synthesized at 1100°C and 32 kb were found to be $\alpha = 1.593$, $\gamma = 1.597$ (indices ± 0.002), optically positive, $2V = 64^\circ\text{C}$. These optical and X-ray data are in reasonable agreement with those published by other investigators [Coes, 1953; Khitarov, Slutskiy, and Arsen'yeva, 1957] with the exception that Khitarov and co-workers report the coesite synthesized by them to be optically negative.

Experimental data. Runs on which the location of the transition curve is based are given in Table 2 and Figure 1. The listed pressures are load pressures corrected for a friction of 8 per cent. In Figure 1 the light dashed lines

roughly paralleling the curve represent the data corrected for -3 per cent and -13 per cent friction; as previously shown [Boyd and England, 1960], these values represent the uncertainty in estimating the friction. The equation for the curve is $P = 19.5 + 0.0112T$, where P is in kilobars and T is in degrees centigrade.

The runs in Table 2 in which the starting material was quartz or cristobalite were brought to temperature in the quartz field and the pressure was then increased to the value desired; runs starting with coesite were brought to temperature in the coesite field and the pressure was decreased to the value desired. All runs listed in Table 2 were quenched by shutting off the power and letting the run cool under pressure. With this procedure the temperature drops to less than 500°C within 5 seconds.

The products of a number of runs in Table 2 consist of a mixture of quartz and coesite. The

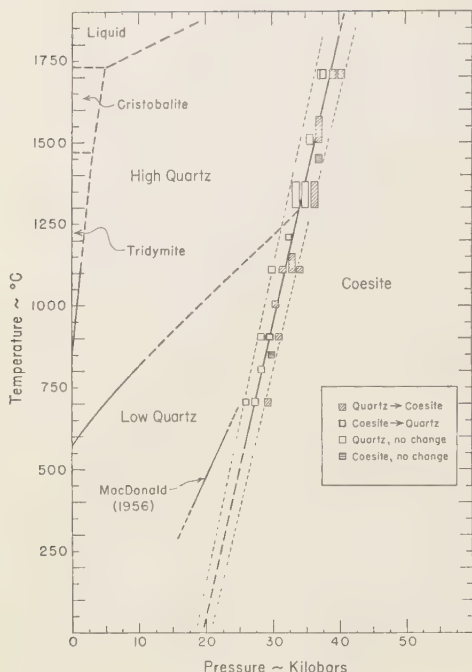


Fig. 1. The quartz-coesite transition with the fields for other polymorphs of SiO_2 . The determination of the quartz-coesite curve by MacDonald [1956] is reproduced. The light, dashed lines on either side of the transition curve represent the uncertainty in correcting the pressure measurements for friction.

percentages of the phases are roughly estimated from peak intensities on the X-ray patterns. Such runs are interpreted as in the coesite field if the starting material was quartz or cristobalite, and as in the quartz field if the starting material was coesite. Since it was found that cristobalite reacts to quartz on the way to temperature, runs with the cristobalite-silicic acid mixes which yielded coesite are shown in Figure 1 as quartz \rightarrow coesite. Runs with the mix coesite + 10 per cent silicic acid yielded either 100 per cent coesite or coesite with definitely more than 10 per cent quartz, and the results are hence not ambiguous.

The reaction has proved to be reversible over the temperature range 700° to 1700°C . The data in Figure 1 and Table 2 shows that the curve obtained by reacting coesite to form

quartz is the same as that obtained by reacting quartz to coesite, within the precision of pressure measurement.

MacDonald's [1956] determination is reproduced in Figure 1. Our data could not be extended into the temperature range investigated by MacDonald because reaction rates are considerably slower in the apparatus used in this study than in the anvil apparatus used by him. MacDonald's curve lies at pressures 10 to 15 per cent less than those given by an extrapolation of our determination. Results obtained with an anvil apparatus similar to that used by MacDonald are described below. They suggest that the effects of pressure gradients across quartz-coesite runs in the anvil apparatus may be responsible for the difference.

Yoder [1950] has determined the high- P quartz inversion in the range up to 10,000 bars. Smooth extrapolation of his curve indicates a triple point high quartz-low-quartz-coesite at approximately 1300°C and 34 kb. There is a detectable break in slope of the quartz-coesite curve at this point.

The fields of tridymite, cristobalite, and liquid are also shown in Figure 1. The boundaries of the fields for tridymite and cristobalite are taken from Tuttle and Bowen [1958, p. 28]. There is no direct experimental determination of the melting curve of quartz, and this slope has been assumed to be $10^\circ/\text{kb}$. The slope of $82^\circ/\text{kb}$ estimated by Mosesman and Pitts [1941] from thermochemical data seems too large in relation to presently available data on other silicate melting curves [e.g. $13^\circ/\text{kb}$ for diopside, Yoder, 1952].

Anvil apparatus data. Before construction of the apparatus used for the experiments described above, we obtained data on the quartz-coesite transition with an anvil apparatus. These data are in poor agreement with the determination shown in Figure 1, and they seem to show evidence of pressure gradients across the sample during a run.

The anvil apparatus used for these runs was modified only slightly from the design developed by Griggs and Kennedy [1956]. The details of the construction and use of this type of apparatus are given in their paper. A sample consisting of a few milligrams of powder was placed between two disks of Pt-10 per cent Fe

TABLE 3. Results of Runs at 700°C on the Quartz-Coesite Transition Made with Anvil Apparatus

All runs had silicic acid as starting material and were brought to temperature at 5 to 10 kb.

Load Pressure, kb	Type of Sealing Ring	Results	Distribution of Coesite
14	None	Quartz	...
15	Nickel	Quartz	...
16	None	Quartz + coesite	Edge
16	Nickel	Quartz	...
17	Nickel	Quartz	...
18	None	Quartz + coesite	Center
18	Nickel	Quartz + coesite	Edge
19	Nickel	Quartz + coesite	Center and edge
20	None	Quartz + coesite	Zone midway between center and edge
20	Nickel	Quartz + coesite	edge
20	Platinum	Quartz + coesite	Edge
20	Platinum	Quartz + coesite	Edge
21	Nickel	Quartz + coesite	Uncertain distribution
21	Platinum	Quartz + coesite	Edge
22	None	Quartz + coesite	Center and edge
22	Nickel	Quartz + coesite	Edge
22	Platinum	Quartz + coesite	Edge
23	Nickel	Quartz + coesite	Center and edge
23	Platinum	Quartz + coesite	Edge
24	Nickel	Quartz + coesite	Edge
24	Platinum	Quartz + coesite	Center and edge
25	Nickel	Quartz + coesite	Center and edge
25	Platinum	Quartz + coesite	Center and edge
26	Nickel	Quartz + coesite	Edge
26	Nickel	Quartz + coesite	Center and edge

oil 0.250 in. in diameter and 0.0005 in. thick. Pressure was applied by compressing the sample between the flat faces of two carbide anvils while the assembly was heated with a Nichrome furnace. In some of the runs a sealing ring of nickel or platinum was used; other runs were made with no sealing ring. The presence of a ring or the type of ring did not seem to influence the results materially. Attempts were made to remove the sample after a run without breaking it so that the distribution of quartz and coesite could be studied with a microscope. Because the samples were removed as disks about 0.001 in. thick, they were sometimes broken and the relative positions of the pieces lost; such runs are listed as 'distribution uncertain' in the tables of data.

Results of a number of runs made at 700°C are given in Tables 3 and 4. In the set of runs listed in Table 3, silicic acid was used as the starting material and the runs were brought to

temperature at a pressure of 5 to 10 kb. It was established by immediate quenches that runs brought to temperature in this way crystallize to fine-grained quartz on the way to temperature. When at temperature, the pressure was increased to the desired value and held for 2 hours. The run was then quenched with the pressure off. With this procedure any coesite that appears in a run has necessarily formed from quartz.

In the sequence of runs in Table 3 coesite first appears with increasing pressure at a load pressure of 16 to 18 kb, about 10 kb lower pressure than the transition curve shown in Figure 1. The coesite found in these runs formed crystals up to nearly 1 mm in diameter. Usually the coesite crystals were clustered into patches surrounded by finer-grained quartz. In more than half of the runs these patches of coesite crystals formed near the margins of the sample. There is a general tendency for the

amount of coesite to increase with increasing pressure, but in an irregular manner. Figure 2, a photograph of a run made at 700°C and a load pressure of 24 kb, shows the sort of coesite

TABLE 4. Results of Runs at 700°C on the Quartz-Coesite Transition Made with Anvil Apparatus

All runs had coesite + 25 to 50% silicic acid as starting material and were brought to temperature at 30 kb.

Load Pressure, kb	Type of Sealing Ring	Results	Distribution of Quartz
28	Nickel	Coesite	...
26	Nickel	Quartz + trace coesite	Center
24	Nickel	Coesite + trace quartz	Edge
22	Platinum	Coesite + quartz	Center
22	Platinum	Quartz	Center and edge
20	Nickel	Coesite + quartz	Center
18	Nickel	Coesite + quartz	Center
16	Nickel	Coesite + quartz	Distribution uncertain

crystals obtained and their tendency to grow around the margins of the sample.

The results of a series of runs in which the starting material was a mixture of coesite and silicic acid are given in Table 4. These runs were brought to temperature at 30 kb, and it was established by immediate quenches that the run brought to temperature in this way had crystallized entirely to coesite by the time the temperature reached 700°C. On this sequence of runs the pressure was dropped from 30 kb to the desired value and quenched after 2 hours with the pressure on. Quartz first appeared in this sequence with decreasing pressure at 26 kb; it usually formed coarse grains in the center of the sample and was surrounded by coesite.

The runs listed in Tables 3 and 4 show that the transition is spread over a pressure interval of about 10 kb at 700°C. The only solid solution possible in this system is the substitution of (OH)₄ groups for SiO₄ tetrahedra in the quartz or coesite in a manner analogous to hydrogrossularite. However, the transition was found to be sharp in the later series of runs shown in Figure 1, where the results of runs in

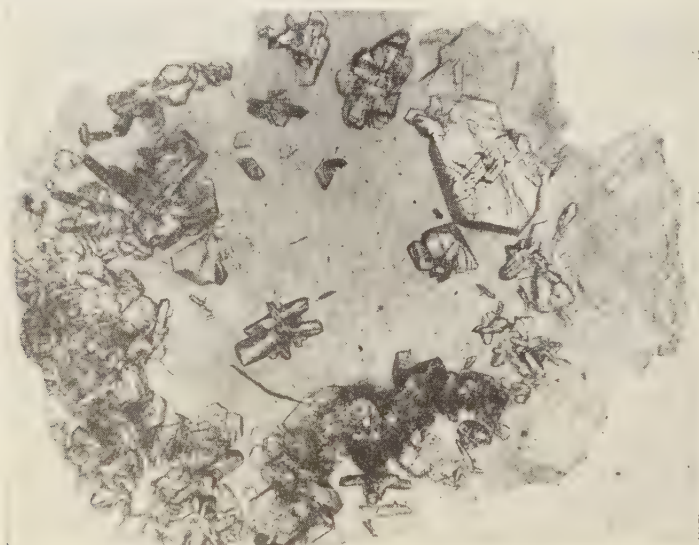


Fig. 2. Run made with anvil apparatus at 700°C and a load pressure of 24 kb. Crystals with high relief are coesite; the matrix is fine-grained quartz. The run is a wafer approximately ¼ in. in diameter in the plane of the photograph and about 0.001 in. thick.

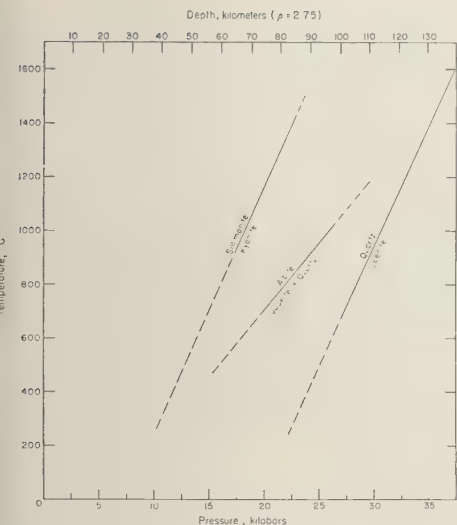


Fig. 3. The quartz-coesite curve compared with the kyanite-sillimanite transition [Clark, Robertson, and Birch, 1957, Fig. 2, curve B] and the reaction albite-jadeite + quartz [Birch and LeComte, 1959].

which H_2O was used as a flux can be seen to be consistent with results obtained dry at higher temperatures. The most probable explanation is that pressure gradients formed across the samples in the anvil apparatus runs. These gradients seem to have developed in an irregular manner, but with a tendency for a higher pressure zone to form around the margins.

In a double anvil or 'squeezer' apparatus the sample itself acts as the pressure medium. It is probable that gradients develop in quartz-coesite runs as a result of the high strengths of these phases. Bridgman [1941] found a compressive strength for quartz of $40,000 \text{ kg/cm}^2$ under a confining pressure of $25,000 \text{ kg/cm}^2$, and coesite is probably stronger than quartz. There is much evidence to suggest that such gradients do not develop in other chemical systems where the reacting phases are more easily sheared and deformed than quartz or coesite.

Geologic discussion. The quartz-coesite boundary is shown in Figure 3 along with two transitions of particular geological interest. The curve for the kyanite-sillimanite transition is from the data of Clark, Robertson, and Birch

[1957]; the reaction albite-jadeite + quartz is taken from the recent work of Birch and LeComte [1959].

The three curves form a sequence consistent with geological observations. Kyanite is abundant in regional metamorphic complexes; the assemblage jadeite + quartz is only rarely found and then only in metamorphic rocks that appear to have formed at low temperatures; coesite has not thus far been found in a natural occurrence. It is probable that regional metamorphism takes place in a crust considerably thicker than that in a nonorogenic area, but the formation of coesite in crustal rocks would require that such rocks be buried to depths of the order of 100 km. It is unlikely that this has happened, but it is possible that coesite will be found in eclogites that have formed at depth in the mantle.

Acknowledgments. The authors wish to thank Drs. H. S. Yoder and Sydney P. Clark and Professor Francis Birch for reading this manuscript and providing many helpful comments.

REFERENCES

- Birch, F., and P. LeComte, Temperature-pressure plane for albite composition, *Am. J. Sci.*, 1959, in press.
- Boyd, F. R., and J. L. England, Apparatus for phase-equilibrium measurements at pressures up to 50 kilobars and temperatures up to 1750°C , *J. Geophys. Research*, 65, 741-748, 1960.
- Bridgman, P. W., Exploration toward the limit of utilizable pressures, *J. Appl. Phys.*, 12, 461-469, 1941.
- Clark, S. P., Jr., E. C. Robertson, and F. Birch, Experimental determination of kyanite-sillimanite equilibrium relations, *Am. J. Sci.*, 255, 628-640, 1957.
- Coes, L., Jr., A new dense crystalline silica, *Science*, 118, 131-132, 1953.
- Griggs, D. T., and G. C. Kennedy, A simple apparatus for high pressures and temperatures, *Am. J. Sci.*, 254, 722-735, 1956.
- Khitarov, N. N., A. B. Slutskiy, and R. V. Arsen'yeva, Synthesis and characteristics of coesite, *Geokhimiya*, 8, 666-672, 1957 (in Russian).
- MacDonald, G. J. F., Quartz-coesite stability relations at high temperatures and pressures, *Am. J. Sci.*, 254, 713-721, 1956.
- Mosesman, M. A., and K. S. Pitzer, Thermodynamic properties of the crystalline forms of silica, *J. Am. Chem. Soc.*, 63, 2348-2356, 1941.
- Tuttle, O. F., and N. L. Bowen, Origin of granite in the light of experimental studies in the sys-

- tem $\text{NaAlSi}_3\text{O}_8$ - KAlSi_3O_8 - SiO_2 - H_2O , *Geol. Soc. Am. Mem.* 74, 1958.
- Yoder, H. S., Jr., High-low quartz inversion up to 10,000 bars, *Trans. Am. Geophys. Union*, 31, 827-835, 1950.
- Yoder, H. S., Jr., Change of melting point of diopside with pressure, *J. Geol.*, 60, 364-374, 1952.
- Zoltai, T., and M. J. Buerger, The crystal structure of coesite, *Z. Krist.*, 111, 129-141, 1959.
- (Manuscript received December 2, 1959.)

Elasticity of Some High-Density Crystals

R. K. VERMA

Dunbar Laboratory, Harvard University, Cambridge, Massachusetts

Abstract. The adiabatic elastic constants of two garnets (spessartite-almandite and almandite), spinel (synthetic), rutile (synthetic), and olivine are reported. The stiffness constants C_{pq} were determined from the velocities of acoustic wave propagation in crystals. The velocities of wave propagation were measured by McSkimin's method. A frequency range of 6 to 12 Mc/s was used.

Introduction. The elasticity of several high-density crystals has been studied. These crystals include two garnets (spessartite-almandite and almandite), spinel (synthetic), rutile (synthetic), and olivine. Some of the common properties of these crystals are low compressibility, high elastic ratio ($\phi_0 = k/\rho$), and high wave velocities (>8.0 km/sec). Such elastic parameters for the compressional waves are attributed to materials below the Mohorovicic discontinuity in the earth [Birch, 1952]. All these crystals have some geophysical importance.

Olivine is of interest because of its relationship to dunites. The wave velocities measured in a , b , and c directions in the crystal give some idea of the range of velocities that may be expected in dunites, which show a considerable anisotropy (Birch, unpublished). The velocities of compressional and shear waves in a quasi-isotropic aggregate of olivine crystals can be calculated from the elastic moduli. These can be compared with the observed velocities in reasonably pure dunites at room temperature and high pressure.

The elastic parameters of the garnets and their wave velocities are of interest in relation to the velocities in eclogites.

Spinel and rutile possess tightly packed crystal structures. It has been suggested [Birch, 1952] that in the mantle under conditions of high pressure common silicates such as olivine and SiO_2 may undergo polymorphic phase transitions to spinel and rutile structures, respectively. The elasticity of spinel and rutile gives some idea of the elastic parameters one might expect as a result of these transitions.

Almost all the stiffness constants C_{pq} of these crystals have been obtained from the velocities

of compressional and shear waves in specific directions in the crystal. The compliances reported have been obtained from these by using appropriate relations.

Notation.

k = Incompressibility

β = Compressibility

μ = Shear modulus

E = Young's modulus

ρ = Density

n_0 = Refractive index

ξ = Direction of particle motion during propagation of a wave

V_p = Velocity of compressional waves

V_s = Velocity of shear waves

ϕ_0 = Elastic ratio = k/ρ

C_{pq} = Stiffness constants

S_{pq} = Compliance constants

Method used for velocity measurement. McSkimin's method [McSkimin, 1950, 1953; McSkimin and Bond, 1957] has been adopted for determining the velocities of propagation of acoustic waves. A block diagram is shown in Figure 1. Briefly, this method involves sending a gated pulse of about 10-Mc/s frequency and 15- to 20- μ sec duration through the specimen which was placed between two fused silica buffer rods. A quarter-wave seal of polystyrene film was used between the specimen and the buffer rods in order to reduce errors due to phase shift between the incident and the reflected pulses at the boundary of the crystal. X- and Y-cut quartz transducers fixed to the buffer rods were used to generate compressional and shear waves. The pulses reflected at the ends of the specimen are picked up, amplified,

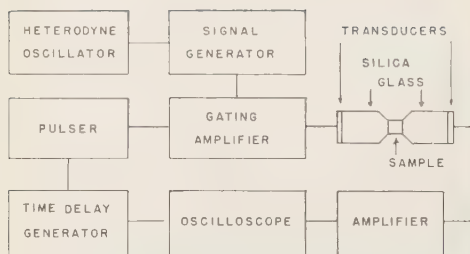


Fig. 1. Block diagram of electrical components.

and displayed on the oscilloscope. If the pulse time is large, then for certain frequencies the reflected pulses arriving in phase interfere constructively and build a stepladder pattern. Figure 2 shows one such pattern for a rutile single crystal. For each in-phase frequency f_0 the velocity of wave propagation is given by the relation [McSkimin, 1950]

$$v = \frac{2tf_0}{n - (\phi/\pi)} \quad (1)$$

where

t = thickness of the crystal in the direction of wave propagation

ϕ = phase shift per reflection at the boundary of the sample

n = number of wavelengths in twice the thickness of the sample

Here n is found from the relation

$$n = f_0/\Delta f \quad (2)$$

where Δf is the frequency difference between two successive in-phase frequencies. As the frequency is varied, a number of in-phase conditions are noted. An average value of f_0 is obtained from these which is used in (2) to calculate n . According to McSkimin, the phase shift is small if a quarter-wave seal for the frequency range is used. In the present work the velocity has been calculated by using the relation

$$v = 2tf_0/n \quad (3)$$

neglecting the phase shift. Velocity is computed for each in-phase condition, and an average of the whole set (15 to 20) is taken as being the representative value.

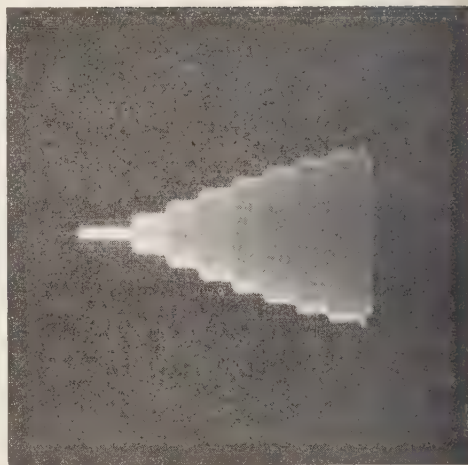


Fig. 2. Reflected pulses in rutile arriving in phase.

This method gives adiabatic values of the stiffness constants from the wave velocities. In most cases the velocities have been measured to an accuracy of about 0.5 per cent. Wherever possible, internal checks have been made on the velocities and the stiffness constants.

Orientation of the single crystals. For orienting single crystals a back reflection Laue camera was used. The principle of this method of orientation is explained by Cullity [1956, p. 215]. A goniometer designed especially for this purpose was borrowed from the Gordon McKay Laboratory of Harvard University. It was possible to orient the crystals in the desired direction to within $1/2^\circ$ or better in most to the cases. The faces perpendicular to the oriented direction were cut on a grinder. The accuracy of parallelism of the faces was about .013 mm.

Theory of wave propagation in crystals. The propagation of plane acoustic waves in a crystalline medium and the derivation of the elastic moduli is treated in various texts [Love, 1944; Cady, 1946; Mason, 1958]. The solution of a wave equation in any direction may be obtained by solving the Christoffel equation [e.g., Mason, 1958, p. 369]. Only the appropriate values of the direction cosines and the stiffness constants of the crystal have to be substituted. It should be noted that for a given direction of wave propagation three modes are possible, each traveling with a characteristic velocity and a

particle displacement direction. Only in a few specific directions are pure compressional and pure shear waves propagated in a crystal. The velocity of propagation of a wave is related to the stiffness constants and the density of the crystal. The number of velocity determinations in a crystal has to be at least equal to the number of independent moduli to be determined.

In the present work the solutions of the wave equation given are according to Voigt's theory unless otherwise stated.

Garnets and spinel belong to the isometric system and have three stiffness constants, C_{11} , C_{12} , and C_{44} . The velocities of compressional and shear waves were determined in [100] and [110] directions. From these the stiffness constants were calculated with well-known solutions of the wave equation in these directions. [e.g., *Mason*, 1958, pp. 368-373].

Rutile belongs to the tetragonal system (class-4/m 2/m 2/m) and has six stiffness constants, C_{11} , C_{33} , C_{44} , C_{66} , C_{12} , C_{13} . Solutions of the wave equation in the [100], [110], and [001] directions were used to determine the first five constants. For determining C_{13} , the solution of the wave equation in a direction inclined at 45° to [100] and [001] was used. The three solutions of the wave equation in this direction are

$$1. \quad \rho V_1^2 = (C_{44} + C_{66})/2$$

$$2. \quad \rho V_2^2 = \frac{C_{11} + C_{33} + 2C_{44} + [(C_{11} - C_{33})^2 + 4(C_{13} + C_{44})^2]^{1/2}}{4}$$

$$3. \quad \rho V_3^2 = \frac{C_{11} + C_{33} + 2C_{44} - [(C_{11} - C_{33})^2 + 4(C_{13} + C_{44})^2]^{1/2}}{4}$$

The second and the third solutions can be used to evaluate C_{13} , but V_2 and V_3 do not correspond to a pure compressional or a pure shear wave velocity unless $C_{11} = C_{33}$. However it is possible to generate experimentally predominantly compressional and shear waves in this direction using X- and Y-cut quartz transducers. The velocity V_2 then corresponds to a predominantly compressional wave velocity and V_3 to a predominantly shear wave velocity, polarized in approximately [101] direction. The two values of C_{13} obtained by this method agree to within about 2 per cent.

Olivine belongs to the orthorhombic system (class, 2/m 2/m 2/m) and has nine constants, C_{11} , C_{22} , C_{33} , C_{44} , C_{55} , C_{66} , C_{12} , C_{13} , C_{23} . The first

six of these have been determined by using compressional and shear waves propagated in the [100], [010], and [001] directions. The last three have been obtained indirectly from Young's moduli ($1/S_{11}$, $1/S_{22}$, $1/S_{33}$) determined by using the composite oscillator method.

Composite oscillator method. This method has been widely used in the study of elastic properties of materials. The theory of this method and the experimental details have been discussed by *Birch* [1950] and other authors cited therein.

For use in this method quartz oscillators of square cross-section were cut with lengths parallel to the Y axis and two faces perpendicular to the X axis. These faces were painted with silver paint. The thickness of these rods was about 2 mm, and the lengths ranged from 5 to 10 mm. The lengths differed by 1.3 mm, so that the successive rods had a frequency difference of about 10 per cent. The bars of the olivine crystal were cut in the axial directions and cemented to the quartz oscillators with shellac flakes. The oscillator was suspended by a spring clip in the plane normal to the axis through the center of the rod, which plane is a node for longitudinal vibrations. Resonance of the oscillator was detected by using an oscillo-

scope across a 0.5 megohm resistance connected in series with the oscillator. The estimated accuracy of resonance detection is 0.2 to 0.3 per cent. The sample was coupled to different quartz oscillators until a composite oscillator was found whose frequency f fell between the frequency of two adjacent oscillators. The following equation was used to find the frequency f_1 of the sample.

$$f_1 = f + (f - f_2)m_2/m_1$$

where

f = frequency of resonance of the composite oscillator

TABLE 1. Chemical Analyses of Garnets and Olivine

	Garnet 1	Garnet 2	Olivine
SiO ₂	35.4%	36.3%	41.2%
Al ₂ O ₃	21.0	21.0	...
FeO	19.1	37.1	8.0
MnO	23.8	0.4	0.1
MgO	0.05	3.6	49.7
CaO	0.44	1.5	0.1
NiO	0.5
Ign.	0.5

f_1 = frequency of resonance of the sample alone

f_2 = frequency of resonance of the quartz crystal

m_1 = mass of the sample

m_2 = mass of the quartz rod

Physical properties of the crystals. Both the garnets (from Brazil) are of gem quality, free of cracks or any major imperfection. Garnet 1 is medium red and garnet 2 is deep red. The spinel and rutile are synthetic crystals purchased from Linde Air Products Ltd. The olivine crystal, of gem quality, is from Burma. The analyses of the garnets and the olivine crystal by J. Ito of the Department of Mineralogy, Harvard University are shown in Table 1. The formulas for the garnets and the olivine were calculated from the above analyses, and those of the spinel and rutile are given by Linde Air Products.

Garnet 1 $3(\text{Mn}_{55}\text{Fe}_{43.5}\text{Mg}_{0.2}\text{Ca}_{1.3})\text{O} \cdot \text{Al}_2\text{O}_3 \cdot 3\text{SiO}_2$

Garnet 2 $3(\text{Fe}_{81}\text{Mg}_{14}\text{Mn}_1\text{Ca}_4)\text{O} \cdot \text{Al}_2\text{O}_3 \cdot 3\text{SiO}_2$

Spinel $\text{MgO} \cdot 3.5(\text{Al}_2\text{O}_3)$

Rutile $\text{TiO}_{1.198103}$

Olivine $2(\text{Mg}_{91.7}\text{Fe}_{8.3})\text{O} \cdot \text{SiO}_2$

In terms of the major end members, garnet 1 consists of a solid solution of spessartite and almandite, and garnet 2 is predominantly almandite. The olivine is predominantly forsterite.

The dimensions of these crystals, as well as other physical properties, are listed in Table 2.

The results. The velocities of wave propagation measured in the [100] and [110] directions in the garnets and in the spinel crystal are given

TABLE 2. Physical Properties and Dimensions of the Crystals

Crystal	Dimensions (mm)	ρ (g/cm ³)	Additional Property
Garnet 1	$t[100] = 9.55$ $t[110] = 8.79$	4.247	$n_o = 1.814$ $a_o = 11.56\text{\AA}$
Garnet 2	$t[100] = 9.60$ $t[110] = 11.35$	4.183	$n_o = 1.817$ $a_o = 11.52\text{\AA}$
Spinel	$t[100] = 13.83$ $t[110] = 12.45$	3.63	
Rutile	$t[100] = 16.19$ $t[001] = 22.45$ $t[110] = 12.18$ $t, 45^\circ \text{ to } [100] \text{ and } [001] = 14.79$	4.264	
Olivine	$t[100] = 16.02$ $t[010] = 11.86$ $t[001] = 9.207$	3.324	

in Table 3. The velocity relation used for each mode of wave propagation is also listed. The elastic moduli obtained from these velocities are tabulated in Table 7, along with those of rutile and olivine.

For rutile the velocities measured in four different directions for various modes are listed in Table 4. It should be noticed that the velocities determining C_{44} are measured in three independent directions. These provide an internal check on the accuracy of the method. The three velocities agree within 0.5 per cent. Also the constants C_{12} and C_{13} have each been determined twice independently.

In the olivine crystal the wave velocities with different modes, measured in the [100], [010], and [001] directions, are reported in Table 5. The stiffnesses C_{11} , C_{22} , C_{33} , C_{44} , C_{55} , and C_{66} were obtained from these whereas C_{12} , C_{13} , and C_{23} were obtained from Young's moduli as explained below.

The data using olivine rods for determining Young's moduli ($1/S_{ii}$) in [100], [010], and [001] directions are shown in Table 6. Correction was applied to the value of v according to the ratio of the lateral dimensions to the length of bar. This correction has been worked out by Bancroft [1941, p. 589]. Here

f_1 = frequency of resonance of the sample

v = velocity of wave propagation in the bar

v_0 = 'bar velocity'

TABLE 3. Wave Velocities Measured in Garnets and Spinel

Crystal	Direction of Wave Propagation	Mode	ξ	Velocity, km/sec	Velocity Relation
Garnet 1	[100]	Compressional	In [100] direction	8.51	$\rho v^2 = C_{11}$
Garnet 2	[100]	"	"	8.54	"
Spinel	[100]	"	"	9.10	"
Garnet 1	[100]	Shear	Any direction in [100] plane	4.74	$\rho v^2 = C_{44}$
Garnet 2	[100]	"	"	4.75	"
Spinel	[100]	"	"	6.61	"
Garnet 1	[100]	Compressional	In [110] direction	8.47	$\rho v^2 = \frac{C_{11} + C_{12} + 2C_{44}}{2}$
Garnet 2	[110]	"	"	8.51	"
Spinel	[110]	"	"	10.30	"

TABLE 4. Wave Velocities Measured in Rutile Crystal

No.	Direction of Wave Propagation	Mode	ξ	Velocity km/sec	Velocity Relation [Voigt, 1928]	Velocity Relation [Raman, 1955]*
1	[001]	Compressional	In [001] direction	10.65	$\rho v^2 = C_{33}$	$\rho v^2 = d_{33}$
2	[100]	"	In [100] direction	9.89	$\rho v^2 = C_{11}$	$\rho v^2 = d_{11}$
3	[001]	Shear	Any direction in [001] plane	5.41	$\rho v^2 = C_{44}$	$\rho v^2 = d_{44}$
4	[100]	"	In [010] direction	3.32	$\rho v^2 = C_{66}$	$\rho v^2 = d_{66}$
5	[100]	"	In [001] direction	5.42	$\rho v^2 = C_{44}$	$\rho v^2 = d_{55}$
6	[110]	"	In [001] direction	5.41	$\rho v^2 = C_{44}$	
7	[110]	"	In [110] direction	6.74	$\rho v^2 = (C_{11} - C_{12})/2$	
8	[110]	Compressional	In [110] direction	8.00	$\rho v^2 = (C_{11} + C_{12} + 2C_{66})/2$	
9	Inclined at 45° to [100] and [001]	"	...	9.98	†	
10	"	Shear	[10 $\bar{1}$]	5.95	‡	

* The significance of these relations will be discussed subsequently.

$$\dagger. \rho v^2 = \frac{C_{11} + C_{33} + 2C_{44} + [(C_{11} - C_{33})^2 + 4(C_{13} + C_{44})^2]^{1/2}}{4}$$

$$\ddagger. \rho v^2 = \frac{C_{11} + C_{33} + 2C_{44} - [(C_{11} - C_{33})^2 + 4(C_{13} + C_{44})^2]^{1/2}}{4}$$

TABLE 5. Wave Velocities Measured in Olivine

No.	Direction of Wave Propagation	Mode	ξ	Velocity km/sec	Velocity Relation [Voigt, 1928]	Velocity Relation [Raman, 1955]*
1	[100]	Compressional	In [100] direction	9.87	$\rho v^2 = C_{11}$	$\rho v^2 = d_{11}$
2	[010]	"	In [010] direction	7.73	$\rho v^2 = C_{22}$	$\rho v^2 = d_{22}$
3	[001]	"	In [001] direction	8.65	$\rho v^2 = C_{33}$	$\rho v^2 = d_{33}$
4	[100]	Shear	In [010] direction	4.88	$\rho v^2 = C_{66}$	$\rho v^2 = d_{99}$
5	[100]	"	In [001] direction	4.87	$\rho v^2 = C_{55}$	$\rho v^2 = d_{66}$
6	[010]	"	In [100] direction	4.88	$\rho v^2 = C_{66}$	$\rho v^2 = d_{88}$
7	[010]	"	In [001] direction	4.42	$\rho v^2 = C_{44}$	$\rho v^2 = d_{55}$
8	[001]	"	In [100] direction	5.00	$\rho v^2 = C_{55}$	$\rho v^2 = d_{77}$
9	[001]	"	In [010] direction	4.54	$\rho v^2 = C_{44}$	$\rho v^2 = d_{44}$

* The significance of these relations will be discussed subsequently.

TABLE 6. Data for Olivine Rods

Direction	Mass, grams	Lateral Dimensions, mm	Length (l), mm	$2l/l = v$, km/sec	$v_0 = (E/\rho)^{1/2}$, km/sec
[100]	.4158	2.858×2.807	16.03	9.33	9.36
[010]	.1823	2.197×2.13	11.86	7.10	7.15
[001]	.1122	2.095×1.78	8.95	7.82	7.91

TABLE 7. Elastic Moduli of Single Crystals
 C_{pq} in units of 10^{11} dynes/cm²; S_{pq} in units 10^{-13} cm²/dyne

	pq	11	22	33	44	55	66	12	13	23
Garnet 1	C_{pq}	30.73			9.52			10.97		
	S_{pq}	4.01			10.50			-1.054		
Garnet 2	C_{pq}	30.48			9.44			11.23		
	S_{pq}	4.09			10.6			-1.102		
Spinel	C_{pq}	30.05			15.86			15.37		
	S_{pq}	5.09			6.31			-1.72		
Rutile	C_{pq}	41.7		48.3	12.5		4.69	2.9	14.7	
	S_{pq}	2.71		2.59	8.0		21.3	0.11	-0.86	
Olivine	C_{pq}	32.4	19.8	24.9	6.67	8.10	7.93	5.9	7.9	7.8
	S_{pq}	3.43	5.88	4.81	14.99	12.4	12.61	-0.67	-0.89	-1.63

The Young's moduli obtained are

$$1/S_{11} = 29.14 \times 10^{11} \text{ dynes/cm}^2$$

$$1/S_{22} = 17.01 \times 10^{11} \text{ dynes/cm}^2$$

$$1/S_{33} = 20.78 \times 10^{11} \text{ dynes/cm}^2$$

For finding the three stiffness constants C_{12} , C_{13} , C_{23} , the relations between the compliances and the stiffnesses were used. These relations are

$$1/S_{11} = C_{11} + \frac{2C_{12}C_{13}C_{23} - C_{12}^2C_{33} - C_{13}^2C_{22}}{C_{22}C_{33} - C_{23}^2}$$

$$1/S_{22} = C_{22} + \frac{2C_{12}C_{13}C_{23} - C_{12}^2C_{33} - C_{23}^2C_{11}}{C_{11}C_{33} - C_{13}^2}$$

$$1/S_{33} = C_{33} + \frac{2C_{12}C_{13}C_{23} - C_{13}^2C_{22} - C_{23}^2C_{11}}{C_{11}C_{22} - C_{12}^2}$$

From these equations we may get a cubic equation in any one of the three unknowns, C_{12}^2 , C_{13}^2 , C_{23}^2 . C_{12}^2 was determined from the cubic equation in C_{12}^2 , which is

$$Z^3 - 690.6Z^2 + 31.76 \times 10^3 Z - 317.3 \times 10^3 = 0$$

where

$$Z = C_{12}^2 \times 10^{-22}$$

The values of the stiffnesses and the compliances obtained are listed in Table 7.

Discussion. Ramachandra Rao [1945] has reported the elastic constants of seven single crystals of garnets without reporting their respective compositions. His values of C_{44} for all garnets (maximum being 8.9×10^{11} dynes/cm²) are lower than our values. Compressibility of his garnet No. 7, reported to be almandite, is 5.2×10^{-13} cm²/dyne. This is much lower than our value of 5.66×10^{-13} cm²/dyne for a garnet of about the same composition. The present value is in close agreement with the value 5.8×10^{-13} cm²/dyne reported by Adams and Gibson [1926] for the compressibility of almandite. Small differences in the values may be expected because of differences in the composition and the quality of the garnets.

Our measurements of the wave velocities in the two garnets suggests that the substitution of Mn for Fe does not make any appreciable

difference in the elastic properties of the garnets belonging to the spessartite-almandite solid solution series.

The present values of the compliance constants for spinel are compared with those obtained by Birch (unpublished) for a similar spinel by using the composite oscillator method. These values are listed below (units of 10^{-13} cm²/dyne).

	Verma	Birch
S_{11}	5.09	5.14
S_{12}	-1.72	-1.80
S_{44}	6.31	6.36

The agreement between the two values is very satisfactory (within ~ 1 per cent).

Bridgman [1928] has reported the compressibility of a natural rutile crystal. The linear and volume compressibilities reported by Bridgman are compared below with the values calculated from the individual elastic moduli of rutile (units of 10^{-13} cm²/dyne).

	Verma	Bridgman
Volume compressibility	4.82	4.83
Linear compressibility		
in c direction	0.88	1.05
in a direction	1.97	1.90

The agreement between the two values is very close, except for the linear compressibility in c direction. This difference may be due to the differences in the composition of the two crystals or imperfections in the natural crystal or both.

The stiffness constants determined for rutile are in wide disagreement with those calculated by Dayal and Appalanarsimham [1950-51]. The observed and the calculated values of these are in units of 10^{11} dynes/cm².

	C_{11}	C_{33}	C_{44}	C_{66}	C_{12}	C_{13}
Calculated	30.0	19	13.3	17.6	17.6	13.6
Observed	41.7	48.3	12.5	4.69	2.9	14.7

The calculated values have been obtained by arguments based upon the crystal structure and on force constants derived from the Raman effect and infrared data [Hearmon, 1956]. It seems that their model of atomic forces may have to be completely revised.

The results obtained for the olivine crystal can be used to calculate the elastic parameters

TABLE 8. Elastic Parameters of Isotropic Aggregate of Olivine

Parameter	According to Voigt [1928]	According to Reuss [1929]	Units
k	13.36	12.88	10^{11} dynes/cm ²
β	7.49	7.76	10^{-13} cm ² /dyne
μ	8.24	7.93	10^{11} dynes/cm ²
V_p	8.56	8.40	km/sec
V_s	4.98	4.89	km/sec

for a quasi-isotropic aggregate of these crystals. The various elastic parameters can be compared with the values experimentally determined for natural aggregates of olivine crystals or dunites. The values of the various parameters according to the theories of Voigt [1928, p. 954] and Reuss [1929] are given in Table 8.

The elastic parameters of the rocks are functions of the pressure, especially in the low pressure range, because of the porosity in the rock. For instance, the compressibility reported by Adams and Gibson [1926] for a dunite from Balsam Gap, N. C., is 8.5 at atmospheric pressure and 7.6×10^{-13} cm²/dyne at 15-kilobar pressure. The present value (according to Reuss's theory) of 7.76×10^{-13} cm²/dyne is close to the latter value.

It is of interest to note the relations between various physical properties of the olivine crystal (Table 9). It is evident that the order of the magnitudes of the wave velocities in various directions is $v_a > v_c > v_b$; this order is opposite that of the linear compressibilities and of the coefficients of thermal expansion. The lowest values of the coefficient of thermal expansion and the linear compressibility should be expected in the direction of closest packing of the atoms, in this case the a direction (and vice versa). The closeness of atomic packing also accounts for the highest velocity of wave propagation in this direction.

Under conditions of higher pressure (constant temperature) the dimensions of the unit cell will change considerably. The b and c dimensions would come much closer to a , the crystal structure thereby tending to approach a higher symmetry. The increase of temperature, at constant pressure has the opposite effect. Under conditions of higher temperature and pressure,

the dimensions of the unit cell would be determined by the dominating factor of these two.

The observations on the rutile and the olivine crystals are sufficient to enable us to make few comments on the generalized theory of elasticity of single crystals proposed by Raman [1955] and Laval [1951]. According to this theory the stress and the strain tensors are assumed to be unsymmetric or $T_{ij} \neq T_{ji}$. The observations on the rutile are particularly reliable because the crystal is free from any noticeable imperfections.

According to Raman, a tetragonal crystal should have nine elastic constants, d_{11} , d_{33} , d_{44} , d_{55} , d_{88} , d_{12} , d_{13} , d_{45} , d_{89} . The first five of these can be determined from the observations given in Table 4. According to Raman's theory the

TABLE 9. Physical Properties of Olivine Direction

	a	b	c	Units
Unit cell dimensions	4.755	10.21	5.985	Å
Wave velocity V_p	9.865	7.726	8.653	km/sec
Linear compressibility	1.878	3.587	2.30	10^{-13} cm ² /dyne
Mean coeff. of thermal expansion (20–100°C) (Rosenholtz, unpublished)	6.78	11.17	9.57	10^{-6} /degree

velocities of No. 3 and No. 5 in Table 4 should have different values in order to reflect the difference in d_{44} and d_{55} . The present results do not support this prediction, as the difference between the two velocities is within the experimental error.

An orthorhombic crystal according to Raman's theory should have 15 stiffness constants. Observations 4 to 9 in Table 5 should correspond to six different constants d_{44} , d_{55} , d_{66} , d_{77} , d_{88} , d_{99} . Thus the values of ρv^2 obtained from observations 4 and 6 should be different, similarly 5 and 8, 7 and 9. Experimentally, we find that the difference between observations 5 and 8 and between 7 and 9 is only 2 to 3 per cent. This difference, although greater than the estimated experimental error, is small. This may be due

TABLE 10. Elastic Parameters of Single Crystals and Their Aggregates According to the Theory of Voigt

Crystal	For Single Crystal		For Aggregates	
	β , 10^{-18} cm ² /dyne	ϕ_0 , (km/sec) ²	V_p , km/sec	V_s , km/sec
Garnet 1	5.69	41.4	8.46	4.77
Garnet 2	5.66	42.2	8.51	4.77
Spinel	4.93	55.8	10.05	5.86
Rutile	4.82	48.5	9.51	5.43
Olivine	7.76	38.7	8.56	4.98

o crystal imperfections or may reflect the order of difference that may be expected between the values of d_{66} and d_{77} , d_{44} and d_{55} .

These high density crystals are of particular interest from the geochemist's point of view. Their common elastic properties are (1) low compressibilities (2) high elastic ratio ϕ_0 , and (3) high wave velocities. For the sake of comparison these parameters are listed in Table 10.

In interpreting the structure of the mantle, Birch [1952] showed that significant changes in the elastic parameters take place in layer C in the mantle. (This layer was defined by Bullen [1953, p. 209] as lying between 413 and 984 km below the earth's crust). These changes of elastic parameters are attributed to changes of composition, phase, or both, in the commonly known silicates; for example, phase change of olivine to spinel structure, or SiO₂ to rutile structure. The values of ϕ_0 determined for olivine and spinel indicate that such a phase change would involve a significant change in the elastic ratio.

Ringwood [1958a, p. 23] has calculated that an 11 per cent increase in the density of olivine may be associated with the phase transition from orthorhombic to isometric (spinel) structure. This would increase the density of olivine from 3.324 to 3.68 g/cm³ on phase transition. Assuming the same incompressibility for olivine in spinel form as for the present spinel (20.26×10^{11} dynes/cm²), we get a value of $\phi_0 = 55.05 \pm 2$ (km/sec)² for this olivine. The incompressibility of our spinel, which has a large stoichiometric excess of Al₂O₃, may be lower than that of Mg₂SiO₄ spinel, for which Ringwood [1958b, p. 206] estimated a value of $\phi_0 = 65 \pm 15$ (km/sec)².

A phase change of SiO₂ to rutile structure would also involve an appreciable change in the elastic ratio (for α -quartz, $\phi_0 = 14$ (km/sec)², Birch [1952, p. 264]).

Velocities of aggregates of rutile and of spinel lie between 9.5 and 10 km/sec. These velocities are found in parts of layer C in the mantle. It is possible to say from the present evidence of the elasticity of these crystals that the elastic properties of layer C would be consistent with phase changes of olivine to spinel structure and of SiO₂ to rutile structure.

Acknowledgments. I am indebted to Professor Francis Birch for his continued guidance, suggestions, and help during the course of this work. My thanks are due Professor C. S. Hurlbut for orienting the olivine crystal and to Mr. H. J. McSkimin for the gating circuit. Kind permission of Professor A. Lang to use the X-ray apparatus is gratefully acknowledged. Skillful preparations of the parts of the apparatus were carried out by Messrs. Harold and Arthur Ames. I am thankful to Professors F. Birch, C. Hurlbut, and N. Bloembergen for reading the manuscript and offering many suggestions.

REFERENCES

- Adams, L. H., and R. E. Gibson, The compressibility of dunite and of basalt glass and their bearing on the composition of the earth, *Proc. Natl. Acad. Sci.*, **12**, 275-283, 1926.
- Bancroft, D., The velocity of longitudinal waves in cylindrical bars, *Phys. Rev.*, **59**, 588-593, 1941.
- Birch, F., A simple technique for the study of elasticity of single crystals, *Am. Mineralogist*, **35**, 644-650, 1950.
- Birch, F., Elasticity and constitution of the earth's interior, *J. Geophys. Research*, **57**, 227-286, 1952.
- Bridgman, P. W., The linear compressibility of thirteen natural crystals, *Am. J. Sci.*, **10**, 287-296, 1928.
- Bullen, K. E., *An Introduction to the Theory of Seismology*, Cambridge University Press, England, 1953.
- Cady, W. G., *Piezoelectricity*, McGraw-Hill Book Co., New York, 1946.
- Cullity, B. C., *Elements of X ray Diffraction*, Addison-Wesley Publishing Co., Reading, Mass., 1956.
- Dayal, B., and N. Appalanarsimham, The evaluation of the elastic constants of rutile from spectroscopic data, *J. Sci. Research, Benares Hindu Univ.*, **1**, 26-30, 1950-51.
- Hearmon, R. F. S., The elastic constants of anisotropic materials, II, *Phil. Mag. Suppl.*, **5**, 323-382, 1956.

- Laval, M. J., Elasticity of crystals, *Compt. rend.*, 232, 1947-1948, 1951.
- Love, A. E. H., *A Treatise on the Mathematical Theory of Elasticity*, Dover, New York, 1944.
- McSkimin, H. J., Ultrasonic measurement technique applicable to small solid specimens, *J. Acoust. Soc. Am.*, 22, 413-418, 1950.
- McSkimin, H. J., Measurements of elastic constants at low temperatures by means of ultrasonic waves, *J. Appl. Phys.*, 24, 988-998, 1953.
- McSkimin, H. J., and W. L. Bond, Elastic moduli of diamond, *Phys. Rev.*, 105, 116-121, 1957.
- Mason, W. P., *Physical Acoustics and the Properties of Solids*, Bell Lab. Series; Van Nostrand Publishers, New York, 1958.
- Ramachandra Rao, B., Elastic constants of garnets, *Proc. Indian Acad. Sci.*, A, 22, 194-198, 1945.
- Raman, C. V., On the theory of elasticity of crystals, *Proc. Indian Acad. Sci.*, A, 42, 51-70, 1955.
- Reuss, A., Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle, *Z. angew. Math. u. Mech.*, 9, 49-58, 1929.
- Ringwood, A. E., The constitution of the mantle-II, *Geochim. et Cosmochim. Acta.*, 15, 18-29, 1958a.
- Ringwood, A. E., The constitution of the mantle-III, *Geochim. et Cosmochim. Acta.*, 15, 195-212, 1958b.
- Voigt, W., *Lehrbuch der Kristallphysik*, B. G. Teubner, Leipzig, 1928.

(Manuscript received October 16, 1959; published under the auspices of the Committee on Experimental Geology and Geophysics and the Division of Geological Sciences, Harvard University; part of a thesis submitted to Harvard University for the Ph.D. degree.)

Letters to the Editor

Changes in the Low-Rigidity Primary Cosmic Radiation
During the Large Forbush Decrease of May 12, 1959F. B. McDONALD¹ AND W. R. WEBBER²*Physics Department
University of Iowa³
Iowa City, Iowa*

One of the largest and most striking Forbush decreases on record occurred on May 12, 1959. The Deep River, Canada, neutron intensity plotted in Figure 1 for the period from April 10 to June 30 shows clearly this large decrease and subsequent exponential recovery (H. Carmichael, private communication, 1959).

Of particular interest are data obtained relating to the rigidity and composition of primary cosmic-ray particles on two balloon flights during this period. The detector used in these experiments consisted of a lucite Cerenkov detector and a NaI scintillation counter. The simultaneous measurement of the Cerenkov counter and the scintillation counter outputs allows a direct determination of the charge and velocity of the incident particles [McDonald and Webber, 1959].

Flight 19 on May 16, 1959, was launched from Minneapolis, Minnesota, and reached an average depth of 6.5 g cm^{-2} at a geomagnetic latitude of 54.4° . This flight occurred near the bottom of the Forbush decrease and gave the lowest intensities we have so far measured for protons and α particles during the recent period of high solar activity. Flight 21 on June 2, 1959, was launched from International Falls, Minnesota, and reached an average depth of 6.5 g cm^{-2} at a geomagnetic latitude of 58.8° . This flight occurred at the conclusion of the exponential

recovery from the Forbush decrease and gave the highest intensities obtained for protons and α particles since September 1956. Because of the particular time sequence of these two flights we believe that the mechanism responsible for the Forbush decrease is also responsible for most, if not all, of the cosmic-ray changes that occurred between May 16 and June 2.

While the measurements of belt 19 on May 16 occurred near the bottom of the Forbush decrease, it is too late to observe the very large increase of low-energy particles reported on May 12, 1959 [Ney, Winckler, and Frier, 1959].

The data from these two flights are summarized in Table 1 and in Figure 2.

The principal features of the data obtained on these two flights are:

1. The rigidity dependence of the Forbush mechanism appears to be similar to that of the

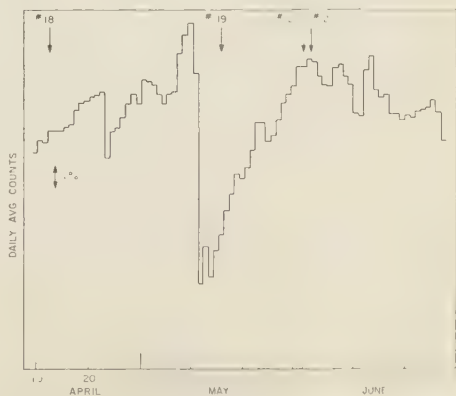


Fig. 1. Daily average counts of Deep River neutron monitor for April, May, and June 1959.

¹ Now at National Aeronautics and Space Administration, Goddard Space Flight Center, Washington 25, D. C.

² Now at Imperial College, London, England.

³ This work supported by joint program of the Office of Naval Research and the U. S. Atomic Energy Commission.

TABLE 1. Integral and Differential Proton and α -Particle Intensities for May 16 and June 2, 1959

Flight Information	Ri- gidity, BV	Kinetic Energy, Mev	No. of Particles/m ² ster sec at 0 Depth with Ri- gidity > R or Kinetic Energy > E	No. of Particles/m ² ster sec BV at 0 Depth with Rigidity R_1 and R_2
A. Protons				
Flt. 19, May 16, 1959	1.12	520	947 \pm 40	0.060 \pm 0.012
avg. = 54.5°	1.42	765	930	
avg. depth = 6.5 g cm ⁻²				
Flt. 21, June 2, 1959	0.61	183	1328 \pm 40	0.036 \pm 0.018
avg. = 59.0°	0.69	230	1325	0.076 \pm 0.024
avg. depth = 6.5 g cm ⁻²	0.78	278	1319	0.120 \pm 0.020
	0.92	373	1303	0.095 \pm - .015
	1.12	520	1287	0.172 \pm 0.020
	1.42	765	1240	
B. α Particles				
	1.08	165	147 \pm 10	0.015 \pm 0.005
	1.40	256	142	0.023 \pm 0.006
Flt. 19	1.68	343	135	0.024 \pm 0.006
	2.03	480	126	0.030 \pm 0.006
	2.58	700	112	
	1.05	155	193 \pm 14	0.023 \pm 0.007
	1.34	240	186	0.024 \pm 0.006
Flt. 21	1.55	305	181	0.029 \pm 0.006
	1.88	420	171	0.039 \pm 0.005
	2.58	700	145	

long-term (11-year variation) [McDonald and Webber, 1959; McDonald, 1959]. For rigidities greater than 2BV this rigidity dependence is similar to the

$$\frac{dJ}{dR} = \frac{dJ_0}{dR} \left(1 - \frac{c(t)}{R} \right)$$

modulation which we have previously suggested.

From May 16 to June 2 Deep River neutrons increased by 8.6 per cent. At the same time α particles with a rigidity (R) > 2.6 BV increased 29 \pm 8 per cent while protons with R > 1.4 BV increased 34 \pm 6 per cent (α particles with R > 1.4 BV increased 32 \pm 8 per cent). In the differential rigidity range from 1.1 BV to 2.6 BV, α particles increased 40 \pm 8 per cent with even more striking changes noted at lower rigidities for the protons. To obtain an extended idea of the rigidity dependence of the changes and correlate them with neutron monitor observations we note that the mean rigidity of particles causing counts in the Deep River neutron monitor is 18 BV. Of more significance is the mean

rigidity of the particles participating in the change. This will be lower than 18 BV since the low-rigidity particles undergo greater changes and will, in fact, depend on the rigidity dependence of these changes. For a $1 - [C(t)/R]$ type of modulation it will be about 8 BV.

The above conclusion (which is also valid for our observation of the March 1956 Forbush decrease) [McDonald, 1957] is not in agreement with the observations of Fenton and others [1958] and McCracken [1959], who find when comparing ratios of changes occurring in neutron monitors and meson telescopes that the long-term decrease appears to have a stronger rigidity dependence than the Forbush-type decrease. Two factors have to be considered, however, when comparing our results with theirs.

(a) A Forbush decrease occurring at times of greatly decreased intensity near solar maximum would not show the same rigidity dependence as observed by sea-level monitors as for the over-all long-term variation, simply because the long-term mechanism has already removed a

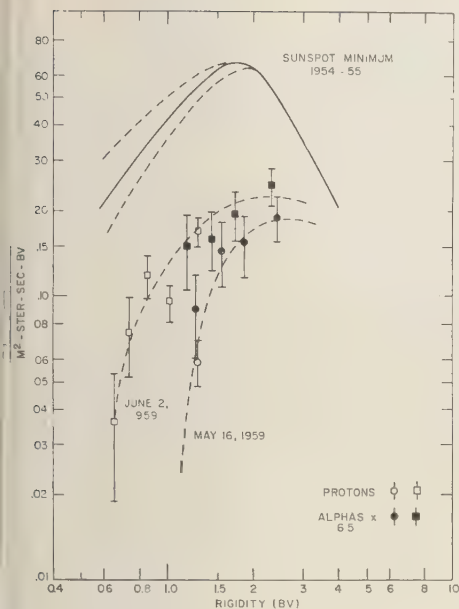


Fig. 2. Low-energy α and proton rigidity spectra. Top solid line is for period of solar minimum. Dotted lines on either side of top solid line represent probable variation during solar minimum.

large part of the low-rigidity particles. The Forbush decrease should then appear less rigidity dependent than the long-term variation even though the form of the modulating mechanism may be very nearly the same. (Such an effect has, in fact, been reported by *McCracken* [1959].)

(b) The neutron monitors and meson telescopes are looking at a different rigidity interval from that viewed in our experiments. Although a neutron monitor at sea level has some sensitivity to particles with rigidities down to 1 BV it derives only 2 per cent of its counting rate from particles with rigidity < 3 BV. This, in effect, means that the rigidity region from 0.6 to 3.0 BV viewed in our experiments is inaccessible to neutron monitors at sea level. This is, of course, just the rigidity interval where the largest changes seem to be occurring and which may provide a crucial test for any new modulation theory. It is entirely possible that the form of the changes occurring may be different in the different rigidity intervals.

2. The protons and α particles appear to participate in a like manner during the Forbush decrease when the data are evaluated as a function of rigidity. The factor of 6.5 which multiplies the α differential and integral spectra to obtain the corresponding proton spectra is found to hold in the Forbush decrease as well as over the long-term variation [*McDonald and Webber*, 1959]. For example, on May 16 the proton- α ratio for $R \ 1.4 > BV$ was 6.55 ± 0.20 , whereas on June 2 it was $6.69 \pm .20$. The quite satisfactory agreement for the differential spectra can be seen in Figure 2.

3. The main features of the Forbush decrease in primary cosmic rays are not due to a change (increase) in cutoff rigidity. The cutoff rigidity at the position of the May 16 flight occurring at the bottom of the Forbush decrease has been previously measured as 1.1 BV [*McDonald*, 1957]. To account for the decrease observed on May 16 the cutoff would have had to increase to approximately 3 BV. The fact that the cutoff for the May 16 flight is 1.1 BV or less indicates:

(a) If the cutoff remained at the previously measured 1.1 BV we should not expect to see particles below that rigidity; however, the differential rigidity spectrum is dropping so fast at these rigidities that we should expect few particles below 1.1 BV even if the cutoff were much lower.

(b) If the cutoff was lower than 1.1 BV at this time (on May 12 at the time of the great magnetic storm it was apparently considerably less than this [*Ney, Winckler, and Frier*, 1959]), the absence of particles below 1.1 BV indicates that the mechanism responsible for the Forbush decrease was capable of almost completely removing particles of this rigidity. The differential intensities must have been less than 0.015 particle/m² ster csec Mev for protons.

In conclusion, then, our results show that the particular Forbush decrease occurring on May 12, 1959 (as well as one reported earlier in March 1956), was indistinguishable in its effects on low-rigidity cosmic-ray protons and α particles from the effects of the long-term variation which has been monitored in more than 10 flights with identical detectors at high latitudes over the past 4 years. By this we do not mean to imply that the two types of changes are caused by the same mechanisms—only that the

end results of these two mechanisms appear to be the same in the low-rigidity region. The difference in mechanisms may prove to be basically only a difference in scale.

REFERENCES

Fenton, A. G., K. B. Fenton, and D. C. Rose, *Can. J. Phys.*, **36**, 824, 1958.

Ney, E. P., J. R. Winckler, and P. Frier, *Phys. Rev. Letters*, **3**, 183, 1959.
McCracken, K. G., *Phys. Rev.*, **113**, 343 (1959).
McDonald, F. B., *Phys. Rev.*, **107**, 1386, 1957.
McDonald, F. B., *Phys. Rev.*, **116**, 462, 1959.
McDonald, F. B., and W. R. Webber, *Phys. Rev.*, **115**, 194, 1959.

(Received November 25, 1959.)

Measurement of Radiation in the Lower Van Allen Belt¹

FRANCIS E. HOLLY

*Air Force Special Weapons Center
Kirtland AFB, New Mexico*

AND

RICHARD G. JOHNSON

*Lockheed Missiles and Space Research Laboratory
Palo Alto, California*

In the first of a series of experiments, an instrument package of eight Geiger counters, shielded and collimated in various ways, was carried aloft and released by an Atlas ICBM; its trajectory passed into the lower Van Allen radiation belt. Since there are variations in altitude at the isocount contours of the belt at several latitudes, the trajectory achieved its deepest penetration into the radiation belt somewhat after reaching its maximum height of 990 km (Fig. 1). Of the eight counters, two with relatively large area and low shielding were saturated by the high flux at deepest penetration and were used only as aids in determining the orientation of the instrument package. The remaining six counters, which provided the data discussed below, were shielded with five different absorber thicknesses (in milligrams per square centimeter of the material indicated) as follows: A, 1.7 mica; B and C, 30 aluminum (2 counters); D, 150 aluminum; E, 400 aluminum; and F, 2000 brass. The electron energies corresponding to these values of extrapolated range are 30, 160, 460, 1000, and 4000 kev, respectively. Counters A, B, and C were end-window tubes collimated to half-angles of 6° by cylindrical entrance apertures. Counters B and C were identical in all respects (including orientation) except that the window of counter C was in a transverse magnetic field, which turned the effective collimator axis for electrons in the energy region of interest by more than 20°; the effective collimator axis for protons of energy great enough to penetrate the absorber was turned a negligible amount. In addition, the

over-all electron sensitivity of this counter in an isotropic flux was lower because of reduced transmission through the collimator, while the proton sensitivity remained the same.

The velocity vectors of trapped radiation at any point in the lower Van Allen belt have been shown to be confined predominantly to a plane. This, plus the fact that the instrument package rotates or tumbles in flight, permitted the use of data from counters B and C to distinguish between electrons and protons (or other heavy

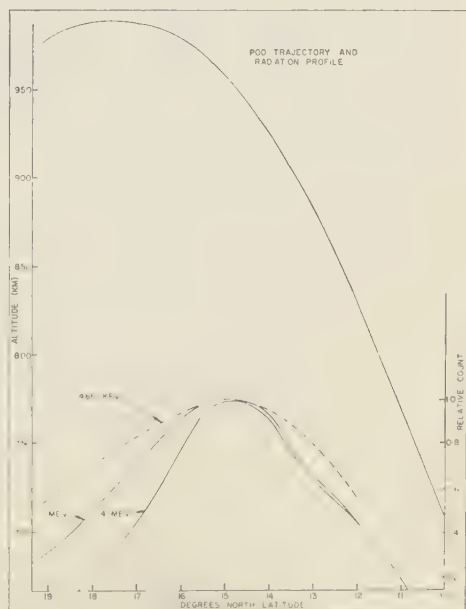


Fig. 1. Variation with altitude of peak counting rate in each counter.

¹ Post deadline paper Washington Meeting, American Physical Society, April 1959.

particles) as follows: If the z axis is taken along the magnetic field, the xy plane contains the collimator axes, separated by more than 20° for electrons but coincident for protons. Two simple cases may be considered. First, if the axis of rotation lies in the plane of the flux and the z axis is parallel to the axis of rotation, each counter will have a peak counting rate twice each revolution as it looks into the plane of the flux. The peaks for the two counters will occur at somewhat different times for electrons but at the same time for protons; also, for electrons the peak rate in counter C will be smaller than in counter B because of the reduced transmission, whereas for protons the peak rates will be equal. Second, if the axis of rotation lies in the plane of the flux and the z axis is perpendicular to the axis of rotation (or a line parallel to it), each counter will again have a peak rate twice each revolution (when the xy plane coincides with the plane of the flux), but the peak rates will occur at the same time in both counters for both electrons and protons; however, the ratio of peak rates in the two counters will still depend on whether the radiation is electrons or protons.

During the deepest penetration of the radiation belt, data were obtained for conditions which at different times closely approximated the idealized cases described. Quantitative analyses yielded an upper limit of 3 per cent (with an uncertainty of 5 per cent) for the proton contribution to the radiation which penetrates 30 mg/cm^2 of aluminum. The analysis of data of the first type described above was the more accurate. Calibrations and necessary corrections were made with the aid of Ti^{504} and W^{185} sources, the spectra of which approximated the observed radiation.

Figure 1 shows the variation with altitude of the peak counting rate in each counter, the curves being normalized to a maximum rate of



Fig. 2. Differential spectrum of observed radiation obtained by subtraction of peak counting rates in successive counters.

unity. Figure 2 shows the differential spectrum of observed radiation, obtained by subtraction of the peak counting rates in successive counters. Less than 1 per cent of the radiation penetrates 150 mg/cm^2 of aluminum. Therefore an estimate can be made of the relative proton contribution to the radiations except in the 150 to 150 mg/cm^2 interval, and this contribution could as well be heavy particles other than protons.

Acknowledgment. The authors wish to acknowledge the assistance of L. Allen, J. A. Welch, Jr., D. J. Knecht, and others of the Air Force Special Weapons Center, and R. D. Moffat, J. Walt, A. Meyerott, and others of Lockheed Missiles and Space Research Laboratory.

(Received December 5, 1959.)

A 'Telescope' for Soft X-Ray Astronomy

RICCARDO GIACCONI

*American Science and Engineering, Inc.
Cambridge, Massachusetts*

AND

BRUNO ROSSI

*Massachusetts Institute of Technology
Cambridge, Massachusetts*

With the development of artificial satellites it has become possible to observe soft X rays from extraterrestrial sources. The purpose of this note is to describe the design of an X-ray telescope' and to analyze some of its characteristics.

The instrument consists of one or several parabolic mirrors on which the X rays impinging at nearly grazing angles undergo total reflection. The possibility of using optics of this type has been discussed in the past in connection with X-ray microscopy [Kirkpatrick and Pattee, 1957; Trurnit, 1946]. These discussions have remained of purely theoretical interest, owing to the difficulty of constructing sufficiently accurate mirrors of the extremely small physical dimensions required. These difficulties, however, are greatly reduced in the construction of large mirrors.

Let us consider first a narrow section of a parabolic mirror whose plane is at the distance l from the focus of the paraboloid, F (Fig. 1). Rays parallel to the axis are concentrated by the mirror into a point at F . It can be shown that, on a first approximation, a parallel beam of rays, forming a small angle, α , with the axis, are concentrated on a circle in the focal plane whose center is at F and whose radius is $R = l\alpha$. Thus, a detector of radius R in the focal plane will record all rays striking the mirror and forming with the axis angles less than R/l .

In the actual design of the instrument it is necessary to consider two limitations: (1) for each wavelength, and for each material, the angle of the incident rays with the reflecting surface must be smaller than a certain value, θ , so that the reflection coefficient will be of the

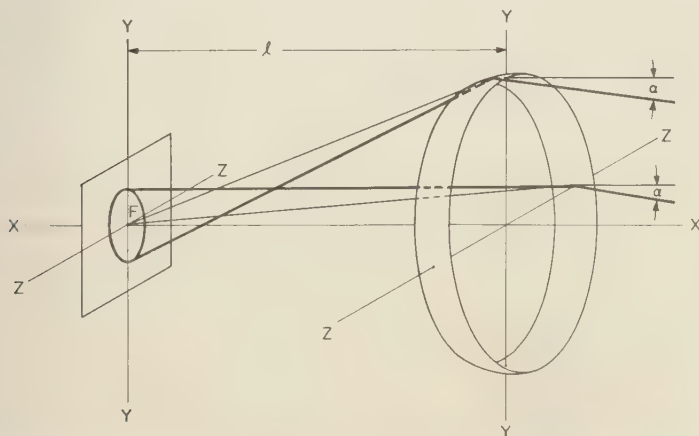


Fig. 1. 'Image' formation by a small segment of a paraboloid. The incident rays are in the xy plane.

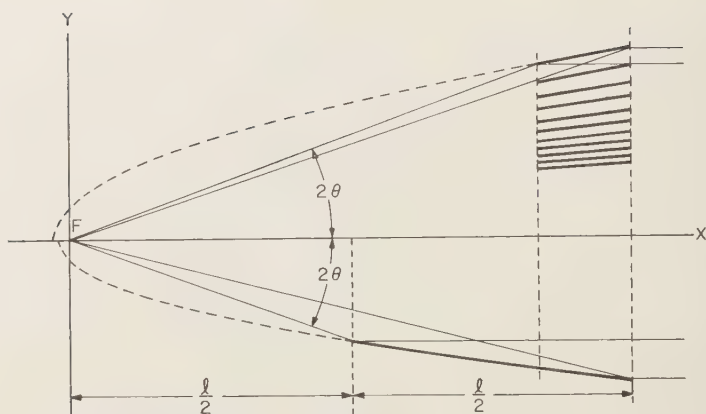


Fig. 2. Lower, the single paraboloid mirror; upper, the system of paraboloid mirrors.

order of unity; (2) in general, the design of the satellite will impose an upper limit to the distance, l , between the detector and the outer edge of the mirror.

The problem is to obtain the maximum area of collection consistent with these limitations. Figure 2 illustrates two possible solutions.

The lower part of this figure is a section of a single parabolic surface. It can be shown that, in this case, the maximum collecting area is obtained by using a paraboloid with focal length $f \simeq \theta^2 l / 2$. The collecting area is then given by $A \simeq \pi(\theta l)^2$. The image of a source off-axis at a small angle, α , is an annulus of outside radius $R \simeq l\alpha$, and inside radius $r \simeq l\alpha/2$.

TABLE 1

The values shown have been computed for silver reflectors with $\theta = 2^\circ$. A and G are computed assuming a reflection coefficient of 1. The actual coefficient is between 0.5 and 1.

Type of Mirror	Single Paraboloid		Set of Paraboloids	
α = resolution, radians	10^{-3}	10^{-4}	10^{-3}	10^{-4}
l = length of the system, cm	200	200	200	200
R = radius of the detector, cm	0.2	0.02	0.2	0.02
A = collecting area of the mirror, cm^2	144	144	576	576
$G = A/\pi R^2$	10^3	10^6	4×10^3	4×10^6

The upper part of the figure shows a section of a mirror composed of segments of confocal paraboloids. The total collecting area, in this case, is approximately $A \simeq \pi(2l\theta)^2$. For an off-axis source, the image is a circle of radius $R \simeq l\alpha$.

For both these systems, two figures of merit may be given: the collecting area of the mirror, A , and the ratio, G , of A to the area of a detector which allows a given resolution α . The value of G is a measure of the signal-to-background ratio; A is a measure of the X-ray gathering power of the mirror.

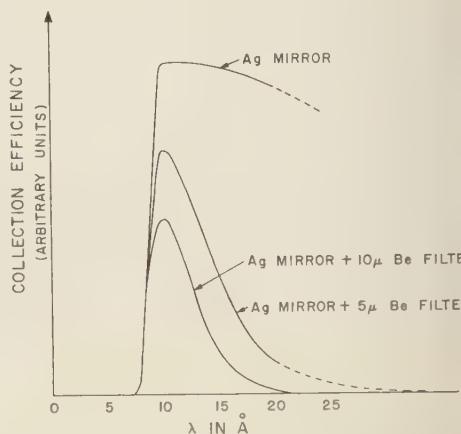


Fig. 3. The dependence of collection efficiency on wavelength for a single silver mirror designed for optimum collection at 10 Å, both with and without the addition of beryllium filters.

Table 1 gives numerical values of these quantities for silver mirrors and for X rays of about 10 Å wavelength. The maximum angle of incidence, θ , has been set equal to 2° , corresponding to a minimum coefficient of reflection of 50 per cent. In Figure 3, the efficiency of light collection for different wavelengths is plotted. Utilizing Table 1, we may estimate the minimum detectable intensity of X rays. The main source of background is cosmic radiation, whose omnidirectional intensity in outer space is of the order of 2 particles/cm² sec. We believe, however, that it is possible to design a detector whose efficiency is 10^3 times higher for X rays than for cosmic-ray particles. Then we see that the minimum detectable intensity is of the order of 10^{-5} quantum/cm² sec for an angular resolution of 10^{-3} radian.

The prime advantages of the instrument are

the large area of collection, the high resolution, and the large signal-to-noise ratio. Among the obvious applications are a detailed analysis of the distribution of X-ray sources on the solar disk and the solar corona, and a search for weak X-ray sources, for example in the Crab Nebula.

We are at present considering the possibility, originally suggested by Wolter [1952] for microscopes, of using multiple total reflections to construct image-forming X-ray telescopes.

REFERENCES

- Kirkpatrick, P., and H. H. Pattee, Jr., X-ray microscopy, S. Flugge, *Encyclopedia of Physics*, 30, 1957.
Trurnit, *Göttinger Nachr., Math. Phys. Kl.*, p. 29, 1946.
Wolter, H., *Ann. Physik*, 10, 94-286, 1952.

(Received December 7, 1959.)

Detection of Sea-Water Motion by Nuclear Precession

E. L. HAHN

*University of California
LaJolla, California**

The long memory of nuclear-spin Larmor precession can be utilized to detect small changes in precession phase angle. The detection of long-range, slow transport of sea water caused by internal waves or other disturbances would seem desirable. Consider how the transport of a volume element of spins in a liquid through a spatial, inhomogeneous, magnetic field affects the phase of Larmor precession. For simplicity, consider a volume element of spins at position x_0 at time $t = 0$, and assume that this volume element moves with a constant velocity v in an inhomogeneous field $H(x)$. The magnitude and direction of v is to be measured, and it is shown that v as small as 10^{-3} cm/sec is detectable. The free precession signal [Hahn, 1950] for this element of spins is defined in terms of the vector

$$V(t) = \exp \left[j\gamma \int_0^t H(x_0 + vt) dt \right]$$

where, at time t , $vt = x - x_0$; γ is the spin gyromagnetic ratio. After writing $H(x_0 + vt)$ as a Taylor's series and keeping the first two terms for convenience, we obtain

$$\begin{aligned} V(t) &= \exp \left\{ j\gamma \int_0^t \left[H(x_0) + \frac{dH}{dx} \Big|_{x_0} vt \right] dt \right\} \\ &= \exp j\gamma [H(x_0)t + Gvt^2/2] \end{aligned}$$

where $G = (dH/dx)_{x_0}$.

We see that the presence of a velocity v produces a phase shift in time t given by $\Delta\varphi = \gamma Gvt^2/2$ (assuming a constant field gradient). In practice, the transport effect can be measured by observing the constructive interference of the various spin volume elements distributed throughout x_0 . Actually x_0 should pertain to

a volume, but an analysis in one dimension is sufficient to give proper orders of magnitude.

Electronic apparatus is necessary here in order to measure spin echoes [Hahn, 1950]. A coil of sufficient volume (a few liters or more, as desired) is immersed and fixed in the sea. As in the Varian magnetometer [Packard and Varian, 1954], protons precess in the earth's field after an initial polarizing field is turned off at $t = 0$. In a time interval from $t = 0$ to $t = \tau$, let the protons precess in a total magnetic field made up of the earth's field H_e , assumed to be perfectly homogeneous, and the inhomogeneous field $H(x)$, supplied by an appropriate secondary coil carrying a current I . At time $t = \tau$, the current I is reversed to $-I$, and $H(x)$ changes to $-H(x)$.¹ Under these conditions, we solve for the spin echo. It is convenient to assume that the fraction of spins which precess in the inhomogeneous field $H(x_0)$ at x_0 is given by

$$P(\Delta\omega) = N \exp(-\Delta\omega^2/2\Delta\omega_s^2)$$

where N is a normalizing coefficient, $\gamma H(x_0) = \Delta\omega$, $\gamma H_e = \omega_s$, and $\Delta\omega_s^2$ is the mean square deviation in Larmor frequency due to $H(x)$. First we compute the phase behavior of $V(t)$ in the time intervals 0 to τ and from τ to ∞ ; we then average $V(t)$ over $P(\Delta\omega)$ and look for constructive interference from spin echoes at some time t_e for $t > \tau$.

At $t = \tau$,

$$V(\tau) = \exp [j(\omega_s + \Delta\omega)\tau + j\gamma Gv\tau^2/2]$$

For $t \geq \tau$, and if we note that $\Delta\omega$ and G change sign (not the G above) because current I is reversed,

¹ Instead of reversing I , a radiofrequency pulse applied for a time t_w at the average Larmor frequency ω_s would serve the same purpose. If the field amplitude of the pulse is H_1 , then $\gamma H_1 t_w = \pi$ is the necessary condition.

* On leave of absence from the Department of Physics, University of California, Berkeley 4, Calif.

$$\begin{aligned}
 V(t) &= \exp(+j\omega_e t) \cdot \exp \left[-j\Delta\omega(t - 2\tau) \right. \\
 &\quad \left. + \frac{j\gamma G v \tau^2}{2} - j\gamma G v \int_{\tau}^t t \, dt \right] \\
 &= \exp \left[j\omega_e t - j\Delta\omega(t - 2\tau) \right. \\
 &\quad \left. - j\gamma G v \left(\frac{t^2}{2} - \tau^2 \right) \right]
 \end{aligned}$$

The measured or average value of $V(t)$ is

$$\begin{aligned}
 \overline{V(t)} &= \int_{-\infty}^{\infty} P(\Delta\omega) V(t, \Delta\omega) \, d\Delta\omega \\
 &= N(2\pi \overline{\Delta\omega_{av}^2})^{1/2} \exp \left[j\omega_e t - j\gamma G v \right. \\
 &\quad \left. \cdot \left(\frac{t^2}{2} - \tau^2 \right) \right] \exp \left[-\frac{(t - 2\tau)^2 \Delta\omega_{av}^2}{2} \right]
 \end{aligned}$$

where $N(2\pi \overline{\Delta\omega_{av}^2})^{1/2} = 1$. A signal maximum occurs essentially at $t = 2\tau = t_*$, and the phase shift of this signal, due to the velocity v , is

$$\Delta\varphi = \gamma G v \tau^2$$

The problem now is to measure $\Delta\varphi$ for a given minimum v to be detected. Assume $G = 0.01$ gauss/cm.²

² To some extent a larger G value could be used but $V(t)$ then attenuates because of molecular self-diffusion [see also Herzog and Hahn, 1956 (appendix)]. From spin-echo experiments, this produces an attenuation of signal amplitude given by

$$\exp[-2/3(\gamma G)^2 D t^2]$$

where D is the self-diffusion coefficient. If we let $\tau = 1$ sec, and $D = 2 \times 10^{-6}$ cm/sec² for water, then attenuation, due to self-diffusion, to 1/e of the initial amplitude, is given when $G \cong 0.01$ gauss/cm. We therefore limit G to this value.

Let $\tau = 1$ sec, which is approximately the relaxation time for sea water, and let $\gamma = 2.7 \times 10^4$ for protons. If we wish to detect a velocity $v = 10^{-3}$ cm/sec, then $\Delta\varphi \cong 0.3$ radians. This phase shift could be measured simply by beating $V(t)$ against a dummy echo signal from an identical apparatus, where $v = 0$. In this way it is possible to cancel out fluctuations in $H(x)$ (or I) which would otherwise cause phase shifts exceeding $\Delta\varphi$. The echo signal will last for a time

$$\Delta t = 2\pi/(\Delta\omega_{av}^2)^{1/2} = 2\pi/\gamma G l \approx 10^{-3} \text{ sec}$$

where $l \cong 30$ cm is chosen as the sample breadth. This should be sufficient time for resolution, since it allows an observed free precession at the average frequency ω_e through at least 10 radians.

If two fixed stations are separated over long distances from one another and each station carries out simultaneous measurements, it would be possible to correlate velocities v at the two positions by noting correlations in phase shifts $\Delta\varphi$. Also the sign of $\Delta\varphi$ would reflect the sign of v . It does not appear feasible to carry out the proposed measurement on a moving platform unless means can be found for correcting or canceling out fluctuations of platform motion to a high degree.

REFERENCES

- Hahn, E. L., Spin echoes, *Phys. Rev.*, **80**, 580-594, 1950.
 Herzog, B., and E. L. Hahn, Transient nuclear induction and double nuclear resonance in solids, *Phys. Rev.*, **103**, 148-166, 1956.
 Packard, M., and R. Varian, Free nuclear induction in the earth's magnetic field, *Phys. Rev.*, **93**, 941, 1954.

(Received September 12, 1959.)

Discussion of Paper by J. D. Isherwood, 'Water-Table Recession in Tile-Drained Land'

E. G. YOUNGS

*Agricultural Research Council Unit of Soil Physics
Huntingdon Road, Cambridge, England*

This paper [Isherwood, 1959] prompts a discussion on the influence of substrata layers on water-table heights in tile-drained land. For steady-state conditions the effect of such layers can easily be determined by means of electric analogue experiments, and there seems to be no reason why the effect for nonsteady-state conditions cannot be inferred from these results. Assurance of this was given by hydraulic experiments in a sand tank in which a drain was installed and on the surface of which rain could be simulated [Collis-George and Youngs, 1958]; the water-table heights observed at the midway position between the drains in nonsteady-state experiments with falling and rising water tables approximated those observed in steady-state conditions for the same drain discharge rate in experiments with different depths of impermeable floor below the drain axis.

Hydraulic and electric analogue experiments (*loc. cit.*) indicated that the effect of an impermeable layer below the drain was negligible for a layer deeper than 0.3 of the half spacing of the drain channels (i.e., 0.15S in Isherwood's notation) below the drain axis. At a depth of 0.25 of the half spacing (or $S/8$) it was very small. However, the effect of the impermeable layer on the water-table height was shown to be dependent on the value of q/K (the ratio of the mean flux of water cutting the water table at any time to the saturated hydraulic conductivity) increasing with decreasing q/K . q was given by the drain discharge divided by the catchment area of the drain and was found to be equal to the rainfall in the corresponding steady state. The smallest value of q/K investigated was 0.01.

Electric analogue experiments were also done for the purpose of investigating the effect on

the water-table height of a very permeable layer below the drain [Youngs, 1959]. Again it was found that the effect was small if the depth of the layer below the drain axis was greater than 0.3 of the half spacing of the drains (0.15S), but in this case the effect was more pronounced for the higher q/K values.

It is convenient to express the variation of the water-table height with the depth of im-

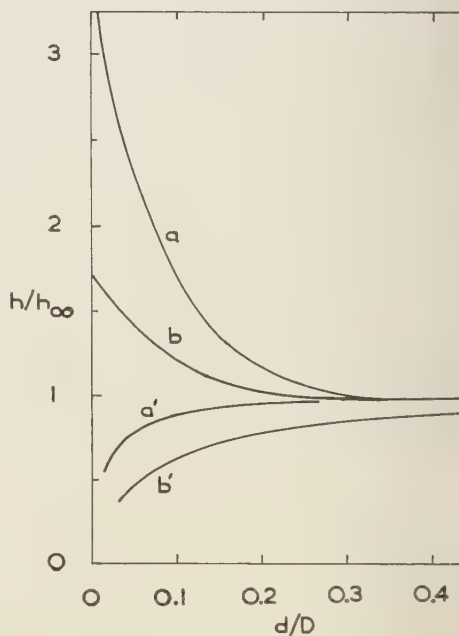


Fig. 1. Variation of the factor h/h_{∞} with d/D [after Youngs, 1959]. (a) $q/K = 0.01$ and (b) $q/K = 0.1$ for a soil overlying an impermeable bed. (a') $q/K = 0.01$ and (b') $q/K = 0.1$ for a soil overlying an infinitely permeable bed.

permeable floor or more permeable bed in terms of the dimensionless ratios h/h_* and d/D , where h is the observed water-table height above the drain axis, h_* that observed for the particular value of q/K if the soil extended to infinite depth, d the depth of the impermeable floor or more permeable bed below the drain axis, and D the half spacing of the drain tubes. In Figure 1 h/h_* is plotted against d/D for q/K values of 0.01 and 0.1 for the two cases of a homogeneous soil overlying an impermeable bed and an infinitely permeable layer. These two cases of substrata layer, the one of zero and the other of infinite hydraulic conductivity, illustrate the extreme effects such layers can have when located below the drain tubes.

Finally, it must be stated that the results quoted here are for drain radii which may be assumed, insofar as the water-table height midway between the drain tubes is concerned, to be running just full of water in the theoretical optimum condition [Childs and Youngs, 1958;

Collis-George and Youngs, 1958; van Deemter, 1950; Engelund, 1951].

REFERENCES

- Childs, E. C., and E. G. Youngs, The nature of the drain channel as a factor in the design of a land-drainage system, *J. Soil Sci.*, 9, 316-331, 1958.
- Collis-George, N., and E. G. Youngs, Some factors determining water-table heights in drained homogeneous soils, *J. Soil. Sci.*, 9, 332-338, 1958.
- Engelund, F., Mathematical discussion of drainage problems, *Trans. Danish Acad. Tech. Sci.*, no. 3, 5-64, 1951.
- Isherwood, J. D., Water-table recession in tile-drained land, *J. Geophys. Research*, 64, 795-804, 1959.
- Van Deemter, J. J., Theoretische en numerieke behandeling van ontwaterings en infiltratiestromingsproblemen, *Verslag. Landbouwk. Onderzoek*, no. 56.7, 1-67, 1950.
- Youngs, E. G., Water-table heights in a drained homogenous soil overlying an infinitely permeable layer, *J. Soil Sci.*, 10, 101-104, 1959.

(Received September 24, 1959.)

The Nonsteady State of the Water Table in Drained Land

E. C. CHILDS

*University of Cambridge School of Agriculture and
Agricultural Research Council Unit of Soil Physics
England*

The interesting papers by *Maasland* [1959] and *Isherwood* [1959] on the nonsteady state of the water table probably develop the discussion of this topic as far as is useful at present. The purpose of this note is to draw attention to an assumption implicit in all work of this kind which needs further examination before progress becomes possible. The assumption is that there is such a quantity as a constant specific yield f , namely the volume of soil water released per unit surface area per unit fall of the water table.

Nonsteady-state moisture profiles. Moisture profiles above a water table which is falling after rainfall cessation may be described, in crude semiquantitative terms only, in some

such manner as proposed in Figure 1 [see *Luthin*, 1957, p. 33]. One may suppose an initial stage indicated by the full line and subsequent stages of which the broken line indicates an example. If one supposes that all such profiles are similar, being merely displaced vertically, then the yield of water, which is transported across the water table, is given by the shaded area between the profiles, and this, in turn, by a simple geometrical argument, is equal to the rectangular area $f dh$ where dh is the fall of the water table and f is the difference between saturation and field capacity, to use a somewhat elusive concept. As so defined the specific yield f is evidently a soil moisture constant. This naive treatment simply ignores and avoids the difficult problem of the water movement and moisture profile development in the unsaturated soil above the water table, a problem in the field of nonlinear diffusion [Luthin, 1957, chap. 1]. It is easy to demonstrate some common cases where such a treatment is greatly in error.

In a period of drainage stress the water table may rise so near to the surface as to restrict the formation of a zone at field capacity (Fig.

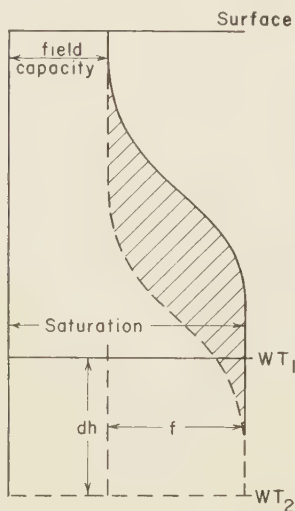


Fig. 1. Hypothetical moisture profiles showing stages of a falling water table in deep soil. The shaded area represents the soil water yield.

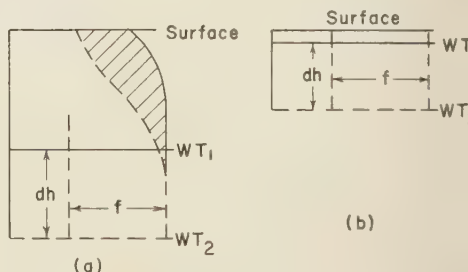


Fig. 2. As for Fig. (1) but (a) shallow water table and (b) very shallow water table. The yield is reduced, to zero in (b).

2a). The yield due to a water table fall dh , even on the same trouble-avoiding basis as that for Figure 1, is still given by the shaded area but is now less than $f dh$, where f is defined as before. In an extreme case (Fig. 2b) where the capillary fringe occupies the whole zone above the water table, the fall dh takes place without any yield of water at all. With increasing lapse of time and fall of the water table the yield will increase until the stage represented in Figure 1 is attained.

Yet again we may have such circumstances as are shown in Figure 3, where the initial profile is appropriate to a fairly intense rate of rainfall. The moisture content above the water table nowhere falls below the value required to endow the soil with a hydraulic conductivity equal to the rate of rainfall, and this may be much in excess of field capacity [Childs, 1945, 1956; Youngs, 1957]. Following cessation of rainfall the moisture profile will tend to become like that of Figure 1 as the water table falls, as shown by the broken line in Figure 3. Here again the water yield is given by the shaded area between the two profiles, but now this is much in excess of the rectangular area $f dh$ and again varies with the stage of water-table fall.

We may summarize by saying that the concept of specific yield as a soil moisture constant is but a rough approximation required to force some preliminary progress; but the true solution of such problems will demand the study of the soil as a whole, both above and below the water table, as an essay in the field of water movement in a medium whose hydraulic conductivity is a function of moisture content.

Errors associated with an assumption of constant specific yield. A highly simplified generalization of the problem, cruder than either Maasland's or Isherwood's, will serve to emphasize both errors and possibilities of the suspect concepts.

Both the earlier examination of drainage problems by electric analogue and the subsequent exact analysis by van Deemter in 1950 and Englund in 1951 [see Luthin, 1957], indicate a very rough proportionality between h , the water-table height midway between neighboring drains, and q , the uniformly distributed flux across the water table. Writing the former as a dimensionless fraction of the drain separa-

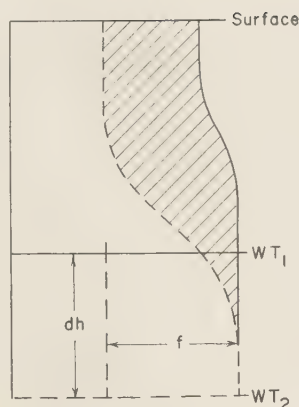


Fig. 3. An initial moisture profile during rainfall and a subsequent profile after cessation of rainfall, showing enhanced yield

tion L and the latter similarly as a fraction of the soil hydraulic conductivity K , we have

$$h/L = b(q/K) \quad (1)$$

where b is a constant. The flux q is the sum of two components, namely the rainfall rate a and the rate of yield due to a falling water table. This latter component may not, in fact, be uniformly distributed, but Collis-George and Youngs [1958] report that equation (1) may nevertheless be used safely if the average yield rate over the water table is taken to be a uniformly distributed flux. Hence we may write

$$q = a - f(dh/dt) \quad (2)$$

Combining (1) and (2), we have

$$f(dh/dt) + (Kh)/(bL) - a = 0 \quad (3)$$

The solution of (3) for constant rainfall rate a is

$$(h - h_u) = (h_0 - h_u) \exp [(-Kt)/(fbL)] \quad (4)$$

where h_0 is the initial water-table height and h_u is the ultimate height given by

$$h_u = abL/K \quad (5)$$

Here we have the exponential decay of the water table demonstrated in certain circumstances by Isherwood [1959]. If the yield is

much smaller than f the decay will be correspondingly more rapid, and if larger will cause a correspondingly delayed decay. Confirmation of this has been provided by experiments, as yet unreported, carried out in the experimental drainage laboratory of the Unit of Soil Physics.

Some years ago it was proposed that steady-state solutions should be applied in practice by equating steady values of rainfall and water-table height to average values over a sufficiently long period of drainage stress [Childs, 1943]. This is justified notwithstanding the variability of yield. Integrating (3) for variability both of a and f gives us

$$\int_{t_0}^{t_1} f dh + (K/bL) \int_{t_0}^{t_1} h dt - \int_{t_0}^{t_1} a dt = 0 \quad (6)$$

Over a sufficiently long time interval of fluctuating rainfall and with the water table oscillating about a mean (in practice a period of a week or two) the first of these integrals may be neglected in comparison with the remaining two, since dh will be as often positive as negative. This amounts to saying that the net change of soil moisture content will be small compared with the total rainfall transmitted to the water table. We may therefore write, dividing throughout by $\int_{t_0}^{t_1} dt$,

$$\left[(K/bL) \int_{t_0}^{t_1} h dt - \int_{t_0}^{t_1} a dt \right] / \int_{t_0}^{t_1} dt = 0$$

$$\text{or} \quad (K\bar{h})/bL - \bar{a} = 0 \quad (7)$$

This is the same as equation (1) with average rainfall rate \bar{a} substituted for steady flux q and average water-table height \bar{h} substituted for the steady value corresponding to q .

REFERENCES

- Childs, E. C., The water table, equipotentials and streamlines in drained land, *Soil Sci.*, 56, 317-330, 1943.
- Childs, E. C., The water table, equipotentials and streamlines in drained land: III, *Soil Sci.*, 56, 405-415, 1945.
- Childs, E. C., Recent advances in the study of water movement in unsaturated soil, *Trans. Intern. Congr. Soil Sci.*, 6th Congr., B, 265-274, 1956.
- Collis-George, N., and E. G. Youngs, Some factors determining water table height in drained homogeneous soil, *J. Soil Sci.*, 9, 332-338, 1953.
- Isherwood, J. D., Water-table recession in tile drained land, *J. Geophys. Research*, 64, 795-804, 1959.
- Luthin, J. N. (Ed.), The drainage of agricultural land. *Agronomy Monograph VII*, Am. Soc. Agron., 1957.
- Maasland, M., Water table fluctuations induced by intermittent recharge, *J. Geophys. Research*, 64, 549-559, 1959.
- Youngs, E. G., Moisture profiles during vertical infiltration, *Soil Sci.*, 84, 283-290, 1957.

(Received October 24, 1959.)

Discussion of Paper by G. W. Smith,
'The Determination of Soil Moisture under a Permanent Grass Cover'

HARRY F. BLANEY

Western Soil and Water Management Research Branch
Soil and Water Conservation Research Division
Agricultural Research Service
U. S. Department of Agriculture
Los Angeles, California

This paper [Smith, 1959] illustrates some interesting comparisons between the use of Thornthwaite, Penman, and Blaney-Criddle formulas to calculate soil-moisture deficits under permanent grass cover as shown in Figures 1 to 6.

Some of the results show a surprisingly close agreement among these formulas, as indicated in Figures 3 and 4. This is especially significant as the coefficients (K) in the *Blaney and Criddle* [1952] formula, $U = KF$ = consumptive use, were based on actual measured evapotranspiration (consumptive use) under normal irrigation practices and climatological observations in arid and semiarid areas rather than potential evapotranspiration as defined by Thornthwaite and Penman. More recently, monthly coefficients (k) have been developed for this formula [Blaney, 1957].

The Blaney-Criddle formula is a modification of the formula $U = K_p F(114 - H)$, developed by *Blaney and Morin* [1942] by correlating evaporation and evapotranspiration with mean temperature, humidity, and average wind velocity (3 miles per hour) and latitude at stations along the Pecos River and Lower Rio Grande valleys in New Mexico and Texas. This formula was used to compute evaporation from free water surfaces and evapotranspiration by phreatophytes in areas where data on water consumption were not available. However, very few data were available on mean humidity in areas where measurements had been made of use of water by irrigated crops.

In a recent trip around the world, the writer found that engineers in Italy, Greece, Turkey, Israel, Pakistan, India, Japan, and Hawaii were using empirical formulas to estimate consumptive water requirements for irrigated crops. In Israel and Japan, research studies are being

conducted on both the Blaney-Criddle and the Thornthwaite formulas.

Smith [1959] mentions the differences of opinion between *Veihmeyer* and others about the influence of soil moisture on the rate of transpiration between field capacity and wilting point. Measurements of actual transpiration use of citrus and avocado trees by *Beckett, Blaney, and Taylor* [1930] in Southern California indicated that soil moisture has very little influence on the rate of transpiration between irrigations as long as moisture is available above the permanent wilting point.

Smith [1959] has developed an ingenious method for comparing actual transpiration with potential transpiration. However, the early studies of Thornthwaite and Penman were made with lysimeters having a high water table, which indicates that moisture was available through capillary action.

REFERENCES

- Beckett, S. H., Harry F. Blaney, and C. A. Taylor, Irrigation water requirement studies of citrus and avocado trees in San Diego County, California, 1926-1927, *Univ. Calif. Agr. Expt. Sta. Bull.* 489, 1-51, 1930.
Blaney, Harry F., and Wayne D. Criddle, Consumptive use of water. A symposium, Paper 2524, *Trans. ASCE*, 117, 949-967, 991-1000, 1952.
Blaney, Harry F., monthly consumptive use of water by irrigated crops and natural vegetation, *Extrait des Compt. rend. rappts. assemblée générale de Toronto, IUGG.*, pp. 432-439, 1957.
Blaney, Harry F., and Karl V. Morin, Evaporation and consumptive use of water empirical formulae, Part I, *Trans. Am. Geophys. Union*, 1942.
Smith, G. W., The determination of soil moisture under a permanent grass cover, *J. Geophys. Research*, 64, 477-483, 1959.

(Received December 14, 1959.)

Magnetic Micropulsations and the Pulsating Aurora

WALLACE H. CAMPBELL

*Geophysical Institute
University of Alaska*

This letter is to report an interesting investigation being undertaken to determine the source of micropulsations in the earth's magnetic field [Campbell, 1959]. At a site 25 miles from College, Alaska, a 2-meter-diameter loop antenna of 21,586 turns is used as a magnetic field sensing unit for pulsations in the range of 1- to 50-sec period. At College, pulsating auroras are measured with lens, interference filter, photomultiplier system.

Figure 1 indicates the 3914 Å auroral pulsations from 70° of sky coincident with magnetic field micropulsations. Similar results for the

5577 Å wavelength have been obtained. A complete report will appear following a thorough study of the phenomenon through the winter of 1958-1959.

The guidance and support of Messrs. A. Belon, C. Deehr, F. Rees, and G. Romick of the Geophysical Institute Optics Division are gratefully acknowledged.

REFERENCE

Campbell, W. H., Studies of magnetic field micropulsations with periods of 5 to 30 seconds. *J. Geophys. Research*, 64, 1819-1826, 1959.

(Received December 18, 1959.)

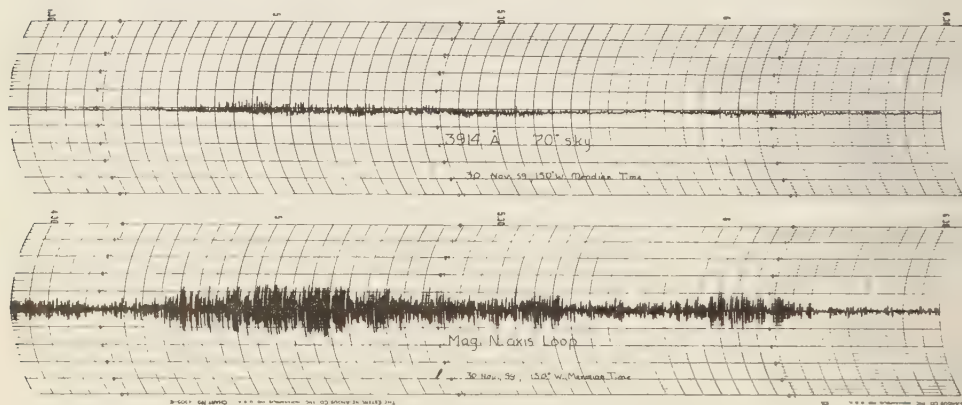


Fig. 1. Micropulsations and Pulsating Aurora.

Discussion of Paper by R. C. Bless, C. W. Gartlein, D. S. Kimball,
and G. Sprague, 'Auroras, Magnetic Bays, and Protons'

PIERRE BERNARD

*Institut de Physique du Globe
Université de Paris
Paris, France*

I have read with interest the paper 'Auroras, magnetic bays, and protons' in the August 1959 issue of your journal, pp. 949-953, and recognized the argument of a magnetic effect of the wind blowing at those altitudes where the electrons, but not the ions, are inhibited to move perpendicular to the magnetic field, an argument I used to explain the 11-year variation of the horizontal component of the geomagnetic field ('Pli cacheté' to the Paris Academy of Sciences, July 13, 1951, and Note in the *Comptes rendus*, 234, 866-868, 1952, which was summarized in *Geophysical Abstracts*, no. 150-13833).

This work received in time the approbation of Father P. Lejay, and Professor Maurain, who

presented the Note, wondered whether it could not furnish an alternative to the dynamo theory.

I am not aware of a previous calculation of the magnetic field of a wind in the transition layer [Bernard, 1955], and I hope, now as 9 years ago, that it may help the understanding of various geomagnetic phenomena.

REFERENCE

Bernard, P., Changements saisonniers dans la variation undécennale de la composante horizontale H du champ magnétique terrestre, *Bull. astron.*, 20, 27-34, 1955.

(Received November 24, 1959.)

A Farside-Type Rocket Experiment for the Measurement of the Gravitational Time Effect

F. W. LIPPS

*Instrument Research, Incorporated
Cambridge, Massachusetts*

The earth's mass causes a gravitational effect on clock rates whose maximum fractional amount is

$$GM/c^2 R_E = 6.96 \times 10^{-10}$$

so that clocks on earth are slow. The effect comes from the Schwarzschild line element of general relativity [Landau and Lifshitz, 1951], which gives the proper time rate in terms of the position and velocity relative to the massive body

$$d\tau/dt = 1 - (GM/c^2 r) - \frac{1}{2}\beta^2$$

Comparison of a clock in a satellite of circular orbit with an earth clock [Singer, 1956; Hoffmann, 1957] should show a fractional change in rate of

$$6.96 \times 10^{-10} [1 - (3/2)(R_E/h)]$$

where h is the height of the orbit from the center of earth. This is mostly a second-order doppler effect for $h < (3/2)R_E$.

To gain the greatest gravitational effect with the greatest economy of rocket energy, a vertically fired rocket may be considered.

Suppose the rocket is launched from the equator and rises to its maximum height h in a time T ; then, neglecting the powered part of the flight, it will return to the launching site if

$$W_E T = \int_0^T dt W(t)$$

where W_E is the earth's angular velocity. Then comparisons can be made between the rocket and a standard clock at the launching site, and the observed average fractional difference in rates will be

$$\frac{\Delta\tau}{2T} = \frac{1}{2T} \int_0^T dt \left\{ \beta_\infty^2 \left(1 - \frac{R_E}{r} \right) + \beta_E^2 - \beta^2 \right\}$$

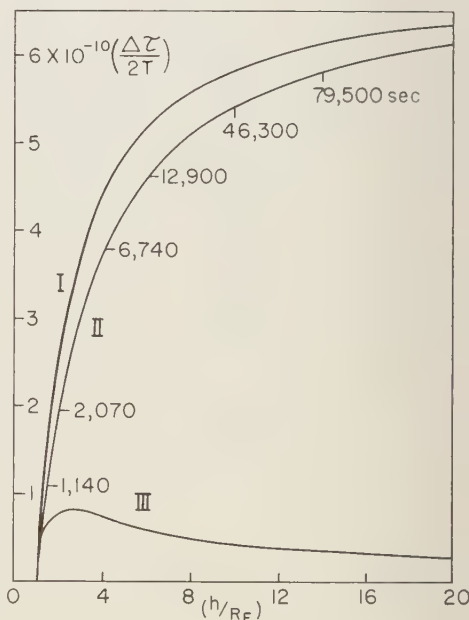


Fig. 1. The result of comparing a clock in a Farside Rocket with a clock on the equator. Curve I is the negative doppler contribution ($-D$), curve III is the gravitational contribution (G), and curve II is the observed fractional effect $\Delta\tau/2T = G + D$. The time of flight to the top of the orbit in seconds is shown on curve II for several heights.

Using the conservation of angular momentum and energy, we get

$$\omega(r) = \omega_0(R_E/r)^2$$

$$\beta^2(r) = \beta_\infty^2(R_E/r - R_E/h) + \beta_i^2(R_E/h)^2$$

The return to the launching site requires that the transverse velocity at 'burn out' be related to earth's equatorial velocity by

$$\beta_E/\beta t = \alpha_2(h, \beta_i)$$

where α_2 is the second-orbit integral.

$$\alpha_n(h, \beta_i) = \left(\int_{R_E}^h \frac{dr}{v_r} \right)^{-1} \int_{R_E}^h \frac{dr}{v_r} \left(\frac{R_E}{r} \right)^n.$$

$\beta_\infty = 3.74 \times 10^{-5}$ is the escape velocity;

$\beta_E = 1.55 \times 10^{-6}$ is the earth's equatorial velocity; and

$$R_E/c\beta_\infty = 568 \text{ sec.}$$

For $h < 10R_E$, $\beta_i^2 \ll \beta_\infty^2$, and then the gravitational and doppler effect (G and D , respectively) are as follows:

$$\frac{\Delta\tau}{2T} = G + D$$

$$G = \frac{1}{2}\beta_\infty^2(1 - \alpha_1) = 6.96 \times 10^{-10}(1 - \alpha_1)$$

$$D = \frac{1}{2}\beta_E^2 + \frac{1}{2}\beta_\infty^2 \left(\frac{R_E}{h} - \alpha_1 \right)$$

$$D = 0.012 \times 10^{-10} + 6.96 \times 10^{-10} \left(\frac{R_E}{h} - \alpha_1 \right)$$

and the time for an ascent is

$$T = \frac{R_E}{c\beta} \propto \sqrt{\frac{h}{R_E}} \left\{ \sqrt{\frac{h}{R_E}} - 1 + \frac{h}{R_E} \cdot \left(\frac{\pi}{2} - \sin^{-1} \sqrt{\frac{R_E}{h}} \right) \right\}$$

(See Fig. 1).

For very great heights, the same formula holds if h is such that $\beta_i = 0$ allows the earth to turn under the rocket orbit for a whole number of days. The maximum effect (6.96×10^{-10}) for the vertical firings does not differ from the maximum for circular orbits, since the velocity effects go to zero for all the nonhyperbolic orbits at great heights.

At 4000 miles above the surface of earth $\Delta\tau/2T \cong 2 \times 10^{-10}$ in an experiment of 4000-sec duration. This is comparable to an experiment with a circular orbit at this height where the fractional change in the rate of time is 1.74×10^{-10} in a long-duration experiment. Although in the case of vertical firing high clock-stability is required for a relatively short term, a correspondingly greater time resolution must be provided in this experiment as compared with the satellite experiment of long duration.

REFERENCES

- Landau, L., and E. Lifshits, *The Classical Theory of Fields* Addison-Wesley Press, p. 272, 1951.
 Singer, S. F., Application of an artificial satellite to the measurement of the general relativistic "red shift," *Phys. Rev.*, 104, 11-14, 1956.
 Hoffmann, B., General relativistic red shift and the artificial satellite, *Phys. Rev.*, 106, 358-359, 1957.

(Received October 3, 1959.)

Geomagnetic and Solar Data

PART 2: Kp, Ap, Ci, AND SELECTED DAYS, FIRST THROUGH THIRD QUARTERS 1959

TABLE 1. Geomagnetic planetary three-hour-range indices Kp, preliminary magnetic character-figures C, average amplitudes Ap (units 2 γ), and final selected days, January to March 1959.

January 1959										February 1959									
E	1	2	3	4	5	6	7	8	Sum	1	2	3	4	5	6	7	8	Sum	
1	0o	0o	0+	1-	1-	0+	0o	0o	2o	3-	3o	2+	3-	3+	3-	2-	1+	20-	
2	0o	0o	0+	2-	2-	1+	1o	1o	7o	2-	4+	4+	4o	4o	3+	3+	4o	29o	
3	2o	2+	2+	2o	2o	1o	2-	0+	14-	4o	3-	2+	4o	4-	4-	3+	5-	28+	
4	1o	1-	2-	2o	2+	3-	2o	1+	14-	5o	4+	3+	4o	5-	4o	5-	5+	35+	
5	2o	1+	1o	4-	4o	4o	5-	6-	26+	4-	5-	4-	4-	4-	4o	3+	4-	30+	
6	5-	4o	3o	3+	4o	4-	4o	3+	30o	4o	3-	2o	4-	3+	3+	4o	3o	26o	
7	4+	4-	3+	3o	2-	2o	4-	4o	26-	3o	3-	2-	2+	3+	2+	3-	1o	19o	
8	4o	3o	2+	3o	2+	3o	4-	4+	26-	1+	2-	2+	3+	3o	3-	3-	4o	20+	
9	5-	2+	2o	3-	4-	6o	5-	6+	32+	5-	5-	5o	3+	3+	3-	1o	2o	27-	
10	6-	5o	5-	5o	5-	4+	5+	4+	39o	2+	1o	0+	1o	1-	1+	1+	3-	11-	
11	5-	2+	2o	1+	2+	2o	3o	4-	21+	1o	4-	4-	6+	5+	5-	4+	4+	33-	
12	4o	2+	1o	2o	3-	3o	3o	3o	21o	4-	5o	3+	4o	3+	3+	3o	4+	29+	
13	2-	2o	3-	3-	2-	2+	3o	1o	17o	4-	2o	1+	3+	4+	4o	4+	4o	27o	
14	0+	1-	3o	2+	2+	1+	2-	3-	14+	3+	4o	4-	4o	6-	5o	3+	2+	31+	
15	2o	4-	3o	2+	1+	1o	1-	1-	15-	3+	6-	6-	5o	4+	4+	4o	2-	34o	
16	1+	0+	2o	4+	4o	4-	4-	4-	23o	5o	6o	5o	3o	5+	6-	6+	6o	42+	
17	3-	4o	3+	3+	4o	4-	3-	2o	26-	6-	5-	5o	3-	2o	2o	2o	1-	25-	
18	4-	4o	3o	3-	3o	4o	3o	3+	27-	1+	1-	1o	1+	1-	2-	1+	1+	9+	
19	2o	3o	3-	3-	2+	2o	3-	1+	19-	3-	4+	3+	3-	3o	2+	2o	1o	21-	
20	1-	0o	1o	1+	1o	1o	1+	2-	8o	0o	1o	2-	2o	1+	1-	1-	1-	8o	
21	0+	0o	1o	1o	1+	1o	1+	0+	6+	1-	0+	3-	2+	2o	1+	2o	1+	13-	
22	1-	2o	3o	2-	2-	2-	2-	0+	14o	3o	4+	3+	4o	2+	3o	2o	3o	25o	
23	3o	2-	1+	3-	3o	1+	1+	1o	15+	4-	4-	4+	4-	3o	1-	1-	0o	2o	
24	2-	2-	2o	1-	1-	1o	0+	0+	8+	0o	0+	1o	2-	2-	1o	1-	2-	8o	
25	1o	1o	1-	4-	4+	4+	4+	3-	22o	3+	5-	6+	6-	6+	7+	5+	4+	43+	
26	1+	4-	2-	4+	6o	4-	1+	0+	22+	5+	5+	6+	5-	5-	5-	4+	4+	40-	
27	0o	2o	3-	3o	3-	2+	2o	2o	17-	3+	4o	3-	3o	4+	6-	5o	4-	32-	
28	3o	2+	2-	2o	2o	2+	3-	2+	18+	5-	5o	3o	4o	5o	6-	5+	5+	38-	
29	1+	2-	3-	3+	3o	3+	4+	4o	24-										
30	3+	3+	2+	3o	2+	2o	2+	0+	19o										
31	3o	4o	4-	3+	3+	2-	2o	2-	21+										
March 1959										Preliminary C, 1959			Average amplitude A ϕ						
E	1	2	3	4	5	6	7	8	Sum	Jan.	Feb.	Mar.	Jan.	Feb.	Mar.	Jan.	Feb.	Mar.	
1	5-	5o	5o	4+	5-	6-	4o	4+	38-	0.0	0.7	1.4	1	11	42				
2	4o	4+	5-	4o	4+	4-	5-	4o	34-	0.1	1.1	1.2	3	23	31				
3	4-	4+	4-	4-	3+	4-	3+	4o	30-	0.2	1.2	1.1	6	22	23				
4	3+	2+	3o	3+	3-	3-	3+	2o	23-	0.5	1.3	0.9	7	36	14				
5	3o	3o	3o	2o	2-	3-	3+	3-	21+	1.3	1.1	0.8	25	24	12				
6	3-	3-	2-	1o	1+	1-	1o	0+	11+	1.2	1.0	0.1	24	18	6				
7	0+	1-	1o	2o	2-	2o	2+	3-	13-	1.1	0.6	0.4	19	11	6				
8	5-	3+	2o	1+	1o	2-	1-	1o	16-	1.1	0.9	0.6	18	12	11				
9	2-	2-	1+	1o	1o	2-	0+	0o	9-	1.5	1.1	0.1	38	23	4				
10	0o	0o	2-	2-	1-	0+	0o	0+	5-	1.6	0.2	0.1	45	6	2				
11	0+	0o	1+	1+	2-	1-	2+	1+	9o	1.0	1.4	0.2	14	36	4				
12	3+	3+	3+	4-	4o	3+	3+	1+	26-	0.9	1.1	1.0	13	24	18				
13	2o	3o	2+	3-	3+	2+	2-	3+	21-	0.5	1.2	0.7	9	21	12				
14	1-	2o	3o	3-	2+	1+	1o	2-	15-	0.5	1.3	0.4	8	30	8				
15	3-	2o	1+	2+	2-	1o	0+	2-	13-	0.3	1.3	0.2	8	37	6				
16	2+	1o	1+	1o	1-	0+	1-	1o	8+	1.2	1.7	0.1	17	61	4				
17	1+	2+	2o	1o	1o	2-	1o	2-	12o	1.0	1.0	0.2	18	24	6				
18	0+	2o	1-	1o	2-	1+	1-	4o	12-	1.0	0.1	0.4	19	4	7				
19	1o	1o	1+	2+	2o	2-	2+	1-	12+	0.4	0.7	0.2	10	12	6				
20	1+	2o	2o	1+	2-	1o	1o	1-	11o	0.1	0.1	0.1	4	4	5				
21	1o	1-	1+	2+	2o	2o	1+	1+	12o	0.1	0.2	0.2	3	6	6				
22	1-	1-	1o	2-	2o	1+	1o	2o	10+	0.4	0.8	0.2	7	17	5				
23	3-	4-	3o	2-	2-	2+	3-	2-	19+	0.5	0.3	0.7	8	15	11				
24	2o	2-	2o	3-	2+	2o	2o	1+	16o	0.1	0.2	0.5	4	4	8				
25	2+	5+	3+	2+	4+	5+	5o	3+	31+	1.3	1.9	1.3	18	69	31				
26	2+	1+	4o	6+	8o	5+	6+	7+	41o	1.2	1.5	1.8	22	48	81				
27	7+	8-	8o	7-	8+	8+	8-	6+	61-	0.6	1.3	1.9	9	30	178				
28	3+	5+	4+	5+	6+	7o	7+	7+	46+	0.5	1.5	1.8	9	44	87				
29	7-	6+	6+	6-	6-	6o	5+	3-	45-	1.2		1.6	17		73				
30	3+	4-	4+	5o	3o	2o	2o	2o	25+	0.7		1.1	11		20				
31	3-	3o	3-	3o	3o	2+	4o	5o	26-	0.8		1.0	14		19				

TABLE 1. (Concluded)—Final Magnetically selected days, January to March 1959

Month	Five quiet days					Ten quiet days										Five disturbed days				
1959																				
January	1	2	20	21	24	1	2	3	4	14	20	21	22	23	24	6	7	9	10	26
February	10	18	20	21	24	1	7	8	10	18	19	20	21	23	24	4	16	25	26	28
March	9	10	11	16	22	6	9	10	11	16	17	19	20	21	22	1	26	27	28	29

TABLE 2. Monthly mean values of C_i , C_p , and A_p , January to March 1959

Index	January 1959	February 1959	March 1959
Mean C_i	0.74	0.95	0.72
Mean C_p	0.67	0.98	0.74
Mean A_p	14	24	24

COMMITTEE ON CHARACTERIZATION OF MAGNETIC DISTURBANCES ,

J. BARTELS, *Chairman*
University
Göttingen, Germany

J. VELDKAMP
Kon. Nederlandsch Meteorologisch Instituut
De Bilt, Holland

TABLE 3. Geomagnetic planetary three-hour-range indices Kp , preliminary magnetic character-figures C , average amplitudes Ap (units 2γ), and final selected days, April to June 1959

April 1959										May 1959									
E	1	2	3	4	5	6	7	8	Sum	1	2	3	4	5	6	7	8	Sum	
1	4-	4o	4-	1o	1o	1-	1o	0+	15+	1+	4-	2o	2+	3-	1+	2o	2o	17+	
2	2o	1o	1+	2+	1+	1+	3-	2+	13-	2o	3-	2+	1o	1o	1-	2-	3-	14o	
3	3+	3o	3-	2+	2+	2o	1-	3-	20o	1o	1o	1o	1+	1+	2-	4-	12o		
4	3-	3-	3-	2-	1+	2-	1+	1o	13+	3o	2-	2+	3o	3-	1+	3o	5o	23o	
5	3-	2-	1o	2-	2-	2-	2-	1-	13-	6+	5+	3o	4o	1o	2o	1o	0+	26o	
6	2o	1o	1+	1+	1+	2o	2+	2-	13o	1-	1-	0+	1o	1o	1+	0+	1-	6o	
7	2o	2-	1o	2o	2+	1+	3+	2+	16o	1+	1-	1o	1+	1o	1o	1o	3o	10+	
8	3-	3+	3+	3o	4-	3+	3o	5+	28-	3+	5+	5+	5o	4-	4o	3-	4o	33+	
9	6o	6+	5-	4o	2-	2+	5+	5-	35o	3+	3-	3-	4-	4-	4-	3-	3-	25o	
10	5-	5o	8-	7-	7-	7o	5-	3-	48-	3-	2-	3-	4-	2o	3-	3o	3o	23-	
11	4-	3+	4-	4o	3o	2+	4-	5o	29-	3o	2+	4o	3o	4-	4-	4-	6o	29+	
12	4-	3o	2o	3-	2-	1-	3-	18-	7o	7-	7-	6-	7-	8+	5+	5-	51o		
13	3-	2o	3o	2-	3-	2o	2+	1+	18-	2-	2+	4-	3-	3o	3o	3+	22+		
14	4-	3o	2+	2-	2o	2o	3o	3o	20-	1+	1+	0+	1+	3-	2-	1-	1o	10o	
15	4o	2-	2o	1o	1+	2o	2o	2-	17-	2+	2o	3-	5-	5o	5+	5+	32o		
16	2+	1+	2-	3-	1o	1o	2-	1+	13o	6o	6-	6-	4-	2+	2-	2o	4+	31+	
17	0+	1o	3-	3-	3-	1+	2-	2-	14o	2-	1+	3-	3o	4-	4o	2+	2+	21o	
18	2+	1-	1-	1+	1o	2-	2-	0+	10-	3+	5+	4o	2+	2+	3+	4-	5o	29-	
19	1-	1-	1-	1+	1+	2o	1-	2o	9+	3+	3o	2-	3o	3o	3-	3o	3o	23-	
20	1-	1+	1o	2-	2o	1+	1o	1-	10-	3o	2-	2+	2+	2-	2+	3+	2+	19o	
21	3-	2+	2+	3o	2o	2+	1-	0+	16-	3o	3-	2+	3-	2-	2+	4-	3+	22-	
22	1-	0+	0+	1-	1o	1o	1-	0+	5-	3+	2+	3+	3o	2o	4-	3o	3o	24-	
23	0o	0+	0+	4o	5-	5o	7-	6+	27+	4-	3o	2+	2+	2-	2+	3-	2+	21-	
24	4-	4o	4+	4o	5-	4+	2+	4o	31+	3+	4o	6-	5o	5-	4o	5-	7+	39-	
25	3o	4-	4o	3-	3o	4o	4-	4-	28-	6-	4-	4-	3-	2-	3o	3-	2-	27+	
26	4-	4o	3-	3o	2o	3o	3o	4+	26-	3+	1o	1-	2o	1+	1+	2+	2-	14-	
27	3o	3-	2o	3o	3o	3o	2-	4+	22-	2-	2-	1-	1-	1o	0-	1-	1o	10-	
28	2-	2o	3-	3o	2+	2-	2+	4+	20o	0+	1o	1-	1-	1o	0+	1-	1+	6o	
29	4o	4-	3o	3o	4+	5o	4o	2+	29+	1-	0+	0+	1-	0+	1-	0+	2-	5o	
30	2o	2o	3o	2o	5-	5o	5-	2o	25+	1o	0+	0+	2+	3+	2o	1-	1-	11-	
31										2+	3+	2+	2+	5o	4-	4o	4+	27+	

June 1959										Preliminary C. 1959			Average amplitude A _p			
E	1	2	3	4	5	6	7	8	Sum	Apr.	May	June	Apr.	May	June	
1	3-	2o	2o	1+	1o	2-	2+	2o	15o	0.6	0.5	0.3	11	9	7	
2	2+	3-	3o	3+	4o	3o	4o	4+	27-	0.3	0.3	1.0	6	7	19	
3	3+	4o	3-	3+	4-	3o	2o	2+	24+	0.6	0.4	1.0	12	7	16	
4	4-	3o	4o	4-	3-	3-	4-	5o	28+	0.2	1.0	1.1	7	16	22	
5	4-	3o	2-	3-	3-	3-	2-	2-	23-	0.1	1.2	0.8	6	29	15	
6	2-	2-	2+	3+	3-	3o	3o	3o	21-	0.2	0.1	0.7	6	3	12	
7	2+	3o	2-	1+	2o	3-	2o	2-	17-	0.6	0.2	0.5	8	6	8	
8	2+	1+	3+	2o	3-	3+	3+	3+	19+	1.1	1.3	0.7	22	33	11	
9	3o	3+	4o	4+	4o	3o	3o	2+	27-	1.5	0.9	1.0	44	16	19	
10	4-	4+	2+	2+	2o	2o	1+	2o	20o	1.9	1.0	0.7	98	14	12	
11	1+	1+	2-	5o	6o	5-	2o	2-	24-	1.3	1.3	1.1	23	26	24	
12	1o	1o	1o	1-	1o	3-	3o	1+	12-	0.5	1.9	0.2	10	108	6	
13	1+	1-	1-	1-	1+	1-	1+	2-	8+	0.5	0.8	0.1	9	14	4	
14	1-	2-	1o	2+	3+	2+	1+	3o	16-	0.7	0.2	0.4	11	5	9	
15	2-	2+	3+	2+	2o	1o	1o	1+	15o	0.6	1.2	0.4	10	33	8	
16	2o	2-	2-	1+	1+	1o	2+	2+	14-	0.2	1.3	0.2	6	36	6	
17	0+	0+	0+	0+	1o	1+	3-	3+	10-	0.4	0.8	0.3	7	13	6	
18	2+	2+	2o	1+	3+	2+	2o	2-	17-	0.2	1.2	0.6	5	24	8	
19	1+	3o	2o	2o	2+	1o	2o	1+	15o	0.2	0.8	1.4	4	14	7	
20	3-	2-	1+	1+	2+	1+	2+	2+	15+	0.1	0.7	0.4	5	10	8	
21	2-	1o	2-	1o	2o	2+	3o	2+	15o	0.4	0.8	0.4	8	13	8	
22	2-	2-	2o	3+	3o	3+	3-	3o	20o	0.0	0.9	0.8	3	15	11	
23	3-	2o	2o	2-	2-	4-	4-	5-	22o	1.5	0.7	0.9	40	12	15	
24	6-	6o	3+	1+	1o	2-	4-	3-	25-	1.4	1.6	1.1	27	52	26	
25	2o	1-	1o	2-	1-	1o	2-	3o	13o	1.1	1.3	0.4	22	25	6	
26	3-	3+	4+	3o	3o	3-	2-	1-	21+	1.1	0.4	0.8	18	7	14	
27	0+	1+	2+	3o	6o	6-	4o	5o	27o	0.9	0.2	1.3	14	5	31	
28	5o	4+	4+	3+	6-	4o	4o	3+	34o	0.9	0.1	1.3	12	3	34	
29	3o	3+	6o	6o	5o	4+	5-	6+	39-	1.2	0.0	1.5	24	3	51	
30	5-	3o	4+	5+	6o	5-	4o	3o	35o	1.1	0.3	1.4	21	6	38	
31											1.1			22		

TABLE 3. (Concluded)—Final Magnetically selected days, April to June 1959

Month	Five quiet days					Ten quiet days										Five disturbed days				
1959																				
April	5	18	19	20	22	2	4	5	6	16	17	18	19	20	22	9	10	23	24	29
May	6	14	27	28	29	2	3	6	7	14	26	27	28	29	30	8	12	15	16	24
June	12	13	16	17	25	1	12	13	15	16	17	19	20	21	25	4	27	28	29	30

TABLE 4. Monthly mean values of Ci , Cp , and Ap , April to June 1959

Index	April 1959	May 1959	June 1959
Mean Ci	0.71	0.79	0.76
Mean Cp	0.70	0.77	0.73
Mean Ap	17	19	15

COMMITTEE ON CHARACTERIZATION OF MAGNETIC DISTURBANCES

J. BARTELS, *Chairman*
University
Göttingen, Germany

J. VELDKAMP
Kon. Nederlandsch Meteorologisch Instituut
De Bilt, Holland

TABLE 5. Geomagnetic planetary three-hour-range indices Kp , preliminary magnetic character-figures C , average amplitudes Ap (units 2γ), and final selected days, July to September 1959.

July 1959										August 1959									
E	1	2	3	4	5	6	7	8	Sum	1	2	3	4	5	6	7	8	Sum	
1	1+	1+	2-	1-	1o	1-	1o	2+	10o	3+	3+	4+	4-	4-	4-	3+	2+	28-	
2	3+	4+	4+	2+	2o	2o	1-	1-	20-	4o	4o	4-	1+	2-	3-	4o	2-	23o	
3	2-	1+	2-	1o	1+	1o	1+	1o	10+	3o	3o	5o	3o	4-	4-	3o	1+	26-	
4	1-	0+	1o	3o	3-	4o	4o	4o	20-	2-	3-	2-	5-	5-	4-	3-	3+	25o	
5	3+	4-	3+	3o	4-	2o	3o	2+	24+	2o	1+	2o	1+	3-	3+	2o	17+		
6	2-	1+	2o	3-	3-	3-	2o	3-	19+	1o	3o	3-	3+	4o	4+	5o	4-	27o	
7	3-	3+	3+	4+	3o	3+	2o	1+	23-	3+	3o	3-	2+	3-	4-	5o	2+	23o	
8	2-	3-	3+	3o	4-	3+	3o	2o	23-	2+	2-	3-	4o	3+	3o	1o	4-	22-	
9	2-	3+	4+	3-	3o	4+	3+	3+	25+	4o	1+	3+	3+	4o	4+	5-	3-	28-	
10	4o	4-	3+	3-	1o	2-	2o	2-	20o	3o	4+	2-	1+	3o	3-	3+	2+	22-	
11	4-	4o	3o	2-	3+	7-	5+	6+	34o	3-	3o	2o	1+	2o	3-	1-	2-	16o	
12	6+	3-	3-	3-	4+	3o	1+	2+	25+	2-	1+	2-	1o	2-	1+	1-	2o	11+	
13	2+	2+	3-	1+	3o	3-	2-	3+	19+	2-	2-	1+	2-	1+	1+	2o	2-	13-	
14	3-	4+	4-	3o	4o	3-	2-	2-	24-	2-	2-	1+	1+	2-	2-	2o	1+	13-	
15	4+	5o	8o	9-	9-	9o	9-	9-	61o	3-	2-	1+	3+	4o	5o	4+	6-	28o	
16	7+	5+	4+	3o	4-	3+	4+	5o	36+	4+	4+	7-	7+	8+	8o	7-	7+	53o	
17	3o	4+	3+	4-	3+	8+	9-	8+	43+	8-	8+	7-	6+	5+	7o	5o	5+	52-	
18	8o	4-	8o	6-	5-	7o	5-	6o	52-	3o	4+	5+	3o	4+	4o	5-	3-	31+	
19	5o	4+	3+	4o	4o	5o	4o	3+	33o	3+	5-	4+	3o	2o	3o	4o	3o	27+	
20	3o	3-	3+	3+	3-	4-	3o	4+	26o	3-	6-	6-	5o	4o	2+	5o	4o	34+	
21	4-	3o	3-	2+	4-	3o	3+	3o	25-	5-	4-	4o	5o	3+	4o	5+	5-	35-	
22	3o	2o	3o	2o	3-	2o	4+	22-	4-	4+	4-	4o	3o	3o	4-	5o	4o	31o	
23	4o	3+	3o	2+	3-	1o	2o	3-	21o	4-	3+	4-	4-	5-	5-	4+	4+	32+	
24	3o	2-	3-	5o	3+	4-	5+	3o	30-	4-	4-	5o	3+	2+	3o	2o	3o	26o	
25	4+	4+	4+	5o	4-	5-	5+	5-	36+	3o	3+	3o	3+	3+	3+	1o	3o	23o	
26	4o	4+	4o	4o	4o	5-	5o	5-	35-	3o	2-	3o	2-	1o	1+	2o	2-	15+	
27	5-	3+	4+	4+	3-	3o	4+	4o	30o	3-	1o	2o	1+	1o	0+	0+	0+	9o	
28	4-	2+	2+	3-	3o	3-	2+	3o	22o	0o	1-	1+	1o	1-	1-	0+	1+	6o	
29	2+	4+	3-	2o	2+	1+	2-	2-	18-	3o	2-	1+	3-	3-	4o	3-	3-	21-	
30	2-	1o	0o	1-	1-	1-	2-	1o	7+	3o	2o	2o	2o	2+	3o	2+	2-	18+	
31	2-	4-	3-	3+	3o	2o	3o	4o	23+	2-	2+	2+	2+	3o	2+	2+	2+	19-	

September 1959										Preliminary C, 1959			Average amplitude A_p		
E	1	2	3	4	5	6	7	8	Sum	July	Aug.	Sep.	July	Aug.	Sep.
1	3+	3o	3+	4-	3o	4o	4+	3-	27+	0.2	1.0	1.1	5	20	20
2	4+	5+	5o	5-	3o	3+	4+	5-	34o	0.8	0.9	1.2	14	16	34
3	3o	3-	3+	4-	3o	4-	4o	7o	30+	0.1	1.0	1.4	5	20	33
4	7-	7+	7o	6+	6o	7o	5+	6-	51+	0.8	1.0	1.8	15	19	103
5	5o	4-	4o	4o	4o	4+	4+	6-	34+	0.9	0.6	1.3	16	9	34
6	6-	4-	4-	3-	2+	2+	1o	1-	22o	0.6	1.1	0.9	11	22	18
7	1+	1+	1+	2o	2+	3+	2-	2-	14+	0.8	1.0	0.4	14	14	7
8	2-	2o	3-	3+	2+	3o	1+	2+	19-	0.8	0.8	0.6	14	14	10
9	2-	1o	1o	2+	1o	2-	1+	2o	12o	0.9	1.1	0.2	18	22	6
10	2-	1+	1o	1+	2-	2o	2-	4-	14+	0.8	0.8	0.4	13	14	8
11	2o	3+	3o	4-	4-	2+	3o	3-	24-	1.4	0.4	0.9	44	8	15
12	5-	3o	3-	3-	2+	3o	2o	1+	22-	1.1	0.2	0.8	24	5	14
13	3+	1o	2o	1+	2-	1o	2+	5+	18o	0.6	0.3	0.7	11	6	14
14	4-	3+	4+	3+	4+	3+	2o	3+	26+	0.9	0.3	1.1	16	6	18
15	2+	2+	3+	3+	4-	3+	5-	4-	25+	2.0	1.2	1.1	236	27	18
16	4+	5-	2+	2o	2+	3-	3-	2+	23+	1.4	2.0	0.9	47	130	16
17	2+	3-	2o	3o	3+	4o	5o	5o	27+	1.9	1.8	1.2	110	114	23
18	3+	3-	3o	5-	4o	4+	4+	5o	31-	2.0	1.2	1.2	119	28	27
19	6o	6o	3+	5o	4+	2+	3-	3-	32+	1.2	1.1	1.2	31	21	36
20	4o	5o	5+	5o	6o	6-	6+	6-	43o	1.0	1.4	1.6	18	38	61
21	6-	7-	7o	5-	6o	7-	6+	5+	48+	0.9	1.3	1.9	16	34	86
22	6+	7+	7o	6-	6-	5o	5+	2+	44o	0.8	1.1	1.7	13	27	73
23	3-	3-	4-	4+	2+	5-	4+	5o	29o	0.7	1.2	1.2	13	28	24
24	4o	4o	4-	3+	5o	3o	3+	3o	29+	1.2	1.1	1.2	28	20	24
25	5o	5-	4o	4+	4o	6-	3+	5-	36-	1.2	0.9	1.2	38	15	37
26	4o	3o	3+	4-	5-	3o	4-	4o	29+	1.1	0.4	1.2	33	8	23
27	3o	3+	4+	4o	4+	4+	3o	3+	28+	1.1	0.2	1.1	25	5	21
28	4o	3o	4+	3+	4-	3-	3+	3-	26+	0.7	0.1	0.9	13	3	18
29	3o	2-	2o	3+	2+	1+	1+	3-	17o	0.5	0.7	0.5	10	13	9
30	2o	1+	2-	3+	3+	3o	3+	3-	21-	0.1	0.6	0.9	4	9	12
31										0.9	0.6		15	9	

TABLE 5. (Concluded)—Final Magnetically selected days July to September 1959

Month	Five quiet days					Ten quiet days										Five disturbed days				
1959																				
July	1	3	13	29	30	1	2	3	4	6	10	13	23	29	30	15	16	17	18	25
August	12	13	14	27	28	5	11	12	13	14	26	27	28	30	31	16	17	20	21	23
September	7	8	9	10	29	7	8	9	10	11	12	13	16	29	30	4	20	21	22	25

TABLE 6. Monthly mean values of Ci , Cp , and Ap , July to September 1959.

Index	July 1959	August 1959	September 1959
Mean Ci	0.95	0.88	1.06
Mean Cp	0.98	0.87	1.06
Mean Ap	32	23	28

COMMITTEE ON CHARACTERIZATION OF MAGNETIC DISTURBANCES

J. BARTELS, *Chairman*
University
Göttingen, Germany

J. VELDKAMP
Kon. Nederlandsch Meteorologisch Instituut
De Bilt, Holland

PROVISIONAL SUNSPOT-NUMBERS FOR
JULY TO SEPTEMBER, 1959

(Dependent on observations at Zurich Observatory
and its stations at Locarno and Arosa)

Day	July	Aug.	Sept.
1	142	177	290
2	118	210	256
3	138	213	202
4	158	225	161
5	136	212	148
6	127	207	144
7	110	179	135
8	131	175	136
9	129	170	157
10	127	155	141
11	133	180	155
12	135	160	170
13	160	125	148
14	180	139	151
15	185	129	168
16	190	151	130
17	193	158	87
18	195	174	100
19	184	182	120
20	160	180	143
21	132	200	132
22	94	200	155
23	113	205	136
24	105	217	155
25	118	212	105
26	134	220	106
27	156	231	92
28	181	274	87
29	182	301	80
30	193	292	76
31	190	284	
Means.....	149.3	198.0	142.2
No. days.....	31	31	30

Mean for quarter: 165.2 (91 days)

M. WALDMEIER

SWISS FEDERAL OBSERVATORY
Zurich, Switzerland

FREDERICKSBURG THREE-HOUR-RANGE
INDICES K FOR JULY TO
SEPTEMBER 1959

[K9 = 500 γ ; scale-values of variometers in
 γ /mm: D = 2.7; H = 2.5; Z = 2.9]

Gr. day	July 1959		August 1959		September 1959	
	Values K	Sum	Values K	Sum	Values K	Sum
1	2220 1022	11	3452 3333	26	3443 2333	25
2	3342 1111	16	4442 2232	23	4555 3334	32
3	1020 1122	9	3362 3341	25	3334 3237	28
4	1012 2345	18	1224 4333	22	5766 5655	45
5	3343 3223	23	1121 2333	16	4343 3345	29
6	1224 3333	21	2444 4234	27	5432 2211	20
7	3444 3321	24	4422 3323	23	1112 2322	14
8	2333 3332	22	2234 3323	22	1233 2212	16
9	2443 3333	25	4144 3343	26	1101 1122	9
10	4431 1322	20	4411 3232	20	1111 2123	12
11	4432 2754	31	2321 2212	15	2323 3233	21
12	6223 3323	24	2210 1022	10	4322 2232	20
13	2242 3224	21	2222 2132	16	3111 1134	15
14	3433 3323	24	2311 0122	12	3342 3323	23
15	3468 9998	56	3212 4555	27	2242 4343	24
16	7442 2234	28	3466 7656	43	4421 2222	19
17	3433 3888	40	6866 4544	43	2223 3355	25
18	7775 4555	45	4543 4333	29	3223 4345	26
19	5433 3444	30	3532 2234	24	5534 3243	29
20	3333 3323	23	2654 3355	33	4465 6454	38
21	3332 3234	23	4444 3345	31	5564 4554	38
22	3242 2324	22	4442 3343	27	6775 4431	37
23	4332 2123	20	4333 5445	31	2343 1334	23
24	2234 3355	27	4453 2234	27	3432 4332	24
25	4444 4345	32	3433 2223	22	5444 3434	31
26	4343 4454	31	3231 0121	13	4333 3333	25
27	4343 3345	29	3020 1111	9	3344 4323	26
28	4222 3233	21	0120 1002	6	4342 3233	24
29	3532 2222	21	3122 2423	19	2113 2012	12
30	2000 0022	6	2222 2333	19	3123 2232	18
31	2434 2225	24	2223 3222	18		

ROBERT E. GEBHARDT
Observer-in-Charge

FREDERICKSBURG MAGNETIC OBSERVATORY
Corbin, Virginia

PRINCIPAL MAGNETIC STORMS

(Advance knowledge of the character of the records at some observatories as regards disturbances)

Observatory (Observer-in-Charge)	Green- wich date	Storm-time		Sudden commencement			C- figure, degree of ac- tivity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K- index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
College (C. J. Beers)	1959	<i>h m</i>	<i>d h</i>		<i>γ</i>	<i>γ</i>	<i>γ</i>					<i>γ</i>	<i>γ</i>	<i>γ</i>
	Jul. 15	08 02	16 04	s.c.*	-14	-337	160	s	15	6,7	9	840	3780	3060
	Jul. 17	16 38	18 21	s.c.*	62	568	136	s	17	6,7	9	730	4400	2320
	Aug. 16	04 ..	17 24	s	16	5	8	370	2180	1700
	Sep. 3	21 58	04 21	s.c.*	94	-458	109	s	04	6	8	280	2200	1020
	Sep. 20	03 47	22 21	s.c.*	11	-114	16	s	21	3	8	330	2050	1330
	Sep. 25	05 15	25 18	ms	25	6	7	267	1330	720
Sitka (M. L. Clevén)	1959													
	Jul. 15	08 00	16 04	s	15	4,5,6,7	9	560	3760	2150
	Jul. 17	16 38	19 09	ssc.*	+12	-70	-40	s	17	7	9	300	1660	900
	Aug. 16	04 05	18 09	ssc.	-4	+60	+10	s	16	5	9	413	1980	1380
									17	2	9			
	Sep. 3	21 59	05 15	ssc.*	+20	-180	-44	s	4	3	9	190	1330	590
	Sep. 20	05 00	22 21	s	20	5	8	260	1410	910
								21	3	8				
								22	3	8				
Vitteveen (D. V. Sabben)	1959													
	Jul. 11	16 26	12 03	s.c.*	-6	+142	-7	ms	11	6	7	30	300	55
	Jul. 15	08 03	16 16	s.c.*	+6	+50	-3	s	15	3,4,5				
										6,7,8	9	170	2225	915
	Jul. 17	16 39	18 20	s.c.	-12	+348	-12	s	17	6	9	75	655	205
	Aug. 16	04 04	17 24	s.c.*	-3	+19	0	ms	16	4,6,8				
									17	1,2	7	60	440	375
	Aug. 20	04 12	20 15	s.c.*	-24	+65	-4	m	20	2,3				
										4,5	5	25	160	40
	Sep. 3	21 59	6 15	s.c.*	-3	+103	+2	ms	3	8	7	40	315	330
	Sep. 20	11 57	22 21	s.c.*	-3	+54	-3	ms	20	7,6	7	55	295	165
									21					
	Sep. 23	11 00	24 01	ms	23	8	6	20	130	65
Fredericksburg (R.E.Gebhardt)	1959													
	Jul. 4	15 xx	05 16	m	4	8	5	20	122	41
	Jul. 11	16 25	12 18	s.c.*	+8	±176	±17	ms	11	6	7	19	256	128
	Jul. 15	08 03	16 12	s.c.	+7	+108	-15	s	15	5,6,7	9	143	1453	1005
	Jul. 17	16 38	21 11	s.c.*	±30	+194	-43	s	17	6,7,8	8	61	794	502
	Jul. 24	04 xx	28 03	m	24	7,8	5	21	161	79
									25	8	5			
									26	7	5			
									27	8	5			
	Aug. 15	12 xx		m	15	6,7,8	5	25	154	80
	Aug. 16	04 03	19 09	s.c.	-1	+25	-3	s	17	2	8	73	307	365
	Aug. 20	04 13	24 13	s.c.	+7	+108	-23	ms	20	2	6	32	234	80
	Sep. 1	04 xx			3	8	7	40	315	675
Sep. 3	21 59	06 19	s.s.c.	+6	+155	+20	ms	4	2	5	22	107	79	
Sep. 17	17 xx	19 18	m	17	7,8					
								18	7,8					
								19	1					
Sep. 20	02 xx	22 21	ms	22	2,3	7	64	226	206	
Fusson (R. L. Viets)	1959													
	Jul. 11	16 25	12 03	s.c.	-3	+83	+8	ms	11	6	6	18	158	37
	Jul. 15	08 02	16 08	s.c.	+3	+125	+9	s	15	6	9	47	810	200
	Jul. 17	16 38	18 18	s.c.*	+13	+51	+4	s	17	6	8	40	337	72
	Aug. 16	04 04	18 09	s.c.	-1	+18	+1	ms	16,17	3,2,3	7	25	255	100
	Aug. 20	04 12	24 10	s.c.	±5	±87	±6	ms	20	2,3	6	22	147	57
	Sep. 3	21 59	06 11	s.c.*	+2	+57	+3	ms	3,4	8,3	7	22	160	60
	Sep. 20	11 58	22 21	s.c.	+2	+16	1	ms	22	3	7	19	199	66

¹Approximate time of ending of storm construed as the time of cessation of reasonably marked disturbance movements in the traces; more specifically, when the K-index measure diminished to 2 or less for a reasonable period.²s.c. = sudden commencement; s.c.* = small initial impulse followed by main impulse (the amplitude in this case is that of the main impulse only, neglecting the initial brief pulse); ... = gradual commencement.³Signs of amplitudes of D and Z to be taken algebraically; D reckoned positive if towards the east and Z reckoned positive if vertically downwards.⁴Storm described by three degrees of activity: m for moderate (when K-index as great as 5); ms for moderately severe (when K = 6 or 7); s for severe (when K = 8 or 9).

GEOMAGNETIC AND SOLAR DATA

PRINCIPAL MAGNETIC STORMS—Continued

Observatory (Observer-in-Charge)	Greenwich date	Storm-time		Sudden commencement			C-figure, degree of activity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K-index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
San Juan (M. Vazquez)	1959 Jul. 15	h m 08 03	d h 16 20	s.c.	'	+53	γ	s	15	6	8	'	γ	2
	Jul. 17	16 39	18 18	s.c.	+1	+98	-45	ms	17	6,7,8	7	21	522	1
	Aug. 16	04 05	17 12	s.c.	+17	-3	ms	16	3,4,5,6	6	17	199	1
	Sep. 3	22 00	05 03	s.c.	+0.3	+30	-10	ms	3	8	7	14	147	1
Honolulu (G. E. Haraden)	1959 Jul. 11	16 26	12 06	s.c.*	-2	+50	+37	ms	11	6	6	4	80	4
	Jul. 15	08 03	16 09	s.c.	-1	+65	+40	s	15	4	8	11	400	3
	Jul. 17	16 39	19 09	s.c.	+4	+85	+55	ms	17	7,8	7	10	210	8
	Aug. 16	04 04	17 15	s.c.	-0.5	+10	+6	ms	18	1				
	Aug. 20	04 13	21 11	s.c.	-1	+32	+22	ms	20	2	7	4	275	3
	Sep. 3	21 59	5 08	s.c.*	+1	+42	+30	ms	3	8	7	2	70	4
	Sep. 20	11 58	22 21	s.c.	+0.5	+13	+7	ms	21	3	6	6	230	3
	Sep. 20	11 58	22 21	s.c.	+0.5	+13	+7	ms	21	3	6	6	105	3
Huancayo (A. Giesecke) (M. Casaverde)	1959 Jul. 11	16 25	12 01	s.s.c.	+3	+107	+11	ms	11	6	6	3	376	2
	Jul. 15	08 03	16 03	s.s.c.	0	+78	+15	s	15	7	9	20	861	11
	Jul. 17	16 38	18 19	s.s.c.	+3	+287	+31	s	17	6	8	15	623	7
	Jul. 24	09 48	25 20	m	24	5,6,7	5	8	250	4
	Aug. 1	11 20	01 20	ms	01	6	6	2	164	1
	Aug. 15	14 07	16 01	m	15	5,6,7	5	7	286	1
	Aug. 16	04 03	17 20	s.s.c.	0	+26	+4	ms	16	6	7	12	428	4
	Aug. 18	06 35	18 20	m	18	5,6	5	5	256	3
	Aug. 20	04 12	20 23	s.s.c.	+1	+94	+16	m	20	1,7	5	5	176	2
	Aug. 23	11 25	23 21	m	23	5,6,7	5	7	264	4
	Sep. 3	21 59	05 07	s.s.c.*	+2	+97	+13	ms	04	6	6	11	420	5
	Sep. 15	11 10	15 21	ms	15	6	6	5	234	3
	Sep. 20	11 57	22 19	s.s.c.	0	+44	+3	ms	20	5	7	13	390	6
	Sep. 25	00 50	25 22	ms	25	6,7	6	6	235	4
Binza (J. Bodson)	1959 Jul. 11	16 26	12 06	s.s.c.	+3	76	8	m	11	5,6	...	6	185	1
	Jul. 15	08 08	16 12	s.s.c.*	+1	67	5	s	15	6	...	16	748	4
	Jul. 15	16 39	18 18	s.s.c.*	+4	98	9	ms	17	8	...	16	218	5
	Aug. 16	04 08	17 12	s.s.c.	+1	12	0	m	16	4	...	16	286	2
	Aug. 20	04 13	20 24	s.s.c.	+1	65	5	m	20	3	...	10	155	5
	Sep. 3	22 00	04 24	s.s.c.	+2	43	5	m	3	8	...	13	121	3
	Sep. 19	10 53	22 24	s.s.c.	+0	9	1	m	21	6	...	11	222	2
	Sep. 19	10 53	22 24	s.s.c.	+0	9	1	m	21	6	...	11	222	2
Port Moresby (J. A. Brooks)	1959 Jul. 15	08 02	16 08	s.s.c.	+50	+30	ms	17	518	18
	Jul. 17	16 38	18 18	s.s.c.	+1	+45	+65	m	18	325	17
	Aug. 16	...	17 13	ms	16	3,5	6+, 7	>10	>180	>8
	Sep. 3	...	05 00	ms
	Sep. 20	03 ..	22 18	ms	22	2,5	6	9	204	10
Elisabeth-ville/Karavia (A. Alexandre)	1959 Jul. 15	08 02	16 08	s.s.c.*	+1	+59	-7	s	15	3,6,7	...	24	656	8
	Aug. 16	04 04	17 24	s.s.c.*	+1	+11	-1	m	16	3,4,5	...	21	321	2
Apia (J. G. Keys)	1959 Jul. 11	16 24	15 08	s.s.c.	+1	+50	-22	m	11	6,8	5	8	106	5
	Jul. 15	08 03	17 16	s.s.c.	0	+42	-18	ms	15	4,6	7	11	473	50
	Jul. 17	16 39	23 13	s.s.c.	+1	+52	-22	ms	17	8				
	Aug. 16	04 04	19 11	s.s.c.	0	+8	-3	ms	18	1	7	11	287	7
	Aug. 20	04 12	24 11	s.s.c.	0	+36	-24	ms	16	3	7	8	286	60
(No record from 3d 19h-4d 03h)														

PRINCIPAL MAGNETIC STORMS—Continued

Observatory (Observer-in-Charge)	Greenwich date	Storm-time		Sudden commencement			C-figure, degree of activity ⁴	Maximal activity on K-scale 0 to 9			Ranges			
		GMT of begin.	GMT of ending ¹	Type ²	Amplitudes ³			Gr. day	Gr. 3-hr. period	K-index	D	H	Z	
					D	H								Z
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Apia (Cont) J. G. Keys)	1959	<i>h m</i>	<i>d h</i>		<i>'</i>	<i>γ</i>	<i>γ</i>				<i>'</i>	<i>γ</i>	<i>γ</i>	
	Sep. 3	14 17	06 13	ms	4	4	6	8	290	50
	Sep. 17	12 ..	23 09	ms	21	3				
									22	1	6	7	225	50
Hermanus A. M. van Wijk)	1959													
	Jul. 11	16 26	12 02	s.s.c.	+4	+23	+24	m	11	6	5	13	112	118
									12	1	5			
	Jul. 15	08 03	16 15	s.s.c.	+3	+25	+22	s	15	3,6	9	84	496	457
	Jul. 17	16 38	19 18	s.s.c.	+4	+83	+56	s	17	8	8	54	186	250
	Jul. 24	09 ..	27 14	m	24	8	8	20	93	79
									26	8	8			
	Aug. 6	09 ..	07 01	m	6	6	5	16	91	70
	Aug. 16	04 04	19 06	s.s.c.	+2	+8	+6	ms	16	4	7	48	251	181
	Aug. 20	04 12	20 15	s.s.c.	+10	+43	+42	ms	20	2	6	25	137	92
				Large initial impulse in D										
	Sep. 3	21 59	06 12	s.s.c.	+2	+39	+24	ms	3	8	6	37	158	123
									4	2,4	6			
	Sep. 13	21 47	14 00	b.p.	m	13	8	5			
	Sep. 17	19 ..	17 23	b.	m	17	7,8	5			
	Sep. 20	02 03	22 20	b.p.	ms	20	7	6	34	146	142
									22	1,2	6			
				Abrupt change at 11h 58m, Sep. 20.										
	Sep. 23	05 ..	27 00	m	24	5	5	22	79	102
									25	8	5			
Gnangara P.M. McGregor)	1959													
	Jul. 11	16 25	12 09	s.s.c*	+2	+43	+24	m	11	6,8	5	15	143	48
									12	1	5			
	Jul. 15	08 03	16 09	s.s.c*	-11	+23	-30	s	15	6,7	9	75	(581)	375
	Jul. 17	16 38	18 10	s.s.c*	+11	+63	+66	s	17	7,8	8	55	244	217
	Aug. 16	13 44	18 00	s.s.c.	(2 oscillations)			ms	16	5,7,8	7	44	228	321
	Sep. 4	22 00	5 08	s.s.c.	+11	+28	+45	ms	4	5	7	33	162	165
	Sep. 20	02 ..	22 21	ms
Foolangi J. Cleary)	1959													
	Jul. 11	16 26	12 01	s.s.c.	-1	+110	+13	ms	11	6,8	6	12	225	37
									12	1				
	Jul. 15	08 04	16 08	s.s.c*	-8	+55	+4	s	15	6,7	9	150	650	410
	Jul. 17	16 39	18 11	s.s.c.	+2	+55	+4	s	17	7	8	70	370	140
	Aug. 15	13 42	18 09	s.s.c.	+1	+4	0	s	16	5	8	65	270	150
	Sep. 3	21 59	05 08	s.s.c*	-6	-95	+2	ms	3	8	6	35	200	120
									4	3,4				
	Sep. 20	02 ..	22 19	ms	21	2	6	30	155	80
									22	3				
Amberley A. L. Cullington)	1959													
	Jul. 11	16 25	12 03	s.s.c*	+8	+97	-32	m	11	6,8	5	18	209	44
									12	1	5			
	Jul. 15	08 04	16 12	s.s.c*	+5	+42	-15	s	15	6	9	68	868	567
	Jul. 17	13 03	19 15	s.s.c.	+0	+7	-2	ms	17	6,7,8	7	55	448	220
									18	1,3	7			
	Aug. 16	04 05	18 09	s.s.c*	-1	+37	+8	ms	16	5	7	53	263	268
	Aug. 17	02 49	18 09	s.s.c*	+1	+18	-8	ms	17	4	7	38	278	236
	Aug. 20	04 15	21 11	s.s.c*	+5	+116	-15	m	20	2,3,7	5	16	145	64
	Sep. 3	22 00	05 03	s.s.c*	+12	+92	-69	ms	3	8	6	22	198	193
									4	3	6			
	Sep. 19	01 00	19 15	s.s.c*	+2	+36	-8	m	4	3	6	11	71	36
	Sep. 20	11 58	22 16	s.s.c*	+1	+26	-5	ms	22	5	6	36	174	107

Information for Contributors to the *Journal of Geophysical Research*

Manuscripts—Send manuscripts to J. A. Peoples, Jr., Department of Geology, University of Kansas Lawrence, Kansas. Manuscripts, including proof copies of figures, should be submitted in triplicate to expedite review and publication. Manuscripts should be in English, typewritten on heavy paper on one side of page only, double spaced (including abstracts and references), with generous margins.

Ample space should be allowed for mathematical expressions, which should be typed or very plainly written by hand. Particular attention should be given to legibility of subscripts and superscripts and to differentiation between capital and lower case letters. Unusual symbols and cumbersome notations should be avoided. Fractional exponents should be used in preference to root signs, and the solidus (/) should be used for fractions wherever its use will save vertical space.

Authors are urged to have their papers critically reviewed by their associates for scientific validity in manner of presentation and use of English before submitting them for publication.

Abstracts—An abstract must accompany each manuscript. It should be a concise but comprehensive condensation of the essential parts of the paper, suitable for separate publication, and adequate for the preparation of general indexes to geophysical literature.

References and footnotes—References should be indicated in the text by the insertion in brackets of the author's name and the year of publication, thus: [Trelease, 1951]. If the author's name is part of the text only the year is bracketed. If there are two or more references citing different papers published in the same year by the same author, distinguish them by the letters a, b, c after the year. At the end of the paper, list all references alphabetically by the authors' names. Include in each entry the following: name of senior author, followed by his initials; names of junior authors, each preceded by his initials; title of paper (or book); title of publication or journal; volume number; inclusive page numbers; year of publication. Abbreviations of journals follow the style used in *Chemical Abstracts*. If in doubt, give the full title of the publication or journal. When a book is cited, add the publisher's name, the city of publication, and the total number of pages. Reference to specific pages may be made in the text if appropriate. Acknowledge unpublished reports and private communications in the text, not as references. Avoid footnotes to the text; use parenthetical sentences instead of footnotes if possible.

Tables and figures—Material suitable to tabular form should be arranged as a table and may be typewritten on a separate page. Tables must be numbered according to their sequence in the text, and each table should have a title. Column headings should be short and self-explanatory; more complete explanation may be given in footnotes to the table. Authors should avoid repeating in the text material which is given in tables or figures.

Figures should be prepared with the column width of this Journal in mind (a scale of two to four times that of the published figure is usually adequate). Lettering and symbols should be large enough to stand reduction and remain legible. Captions should be typed on a separate page, not lettered in the figures. Necessary legends or lettering in the figures should be executed to meet competent drafting standards, not typewritten. If the author cannot arrange for suitable lettering, he may send the drawings with the lettering lightly penciled in or shown on a proof copy, and the lettering will be done at the editorial office.

Line drawings should be in India ink on white paper or tracing cloth. Coordinate paper should be avoided, but, if used, it must be blue-lined and the coordinate lines which are to show must be inked.

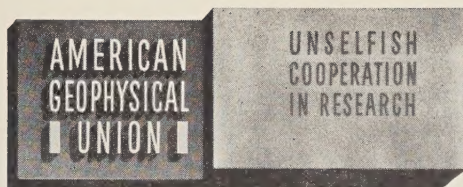
Photographs are acceptable only if they have good intensity and contrast. They should be unmounted, glossy prints.

Figures should be identified by numbering lightly in pencil, and 'top' of each figure should be indicated.

Acknowledgments—Acknowledgments should be made only for significant contributions by the author's professional associates. A brief closing statement will usually suffice.

REFERENCES

- AMERICAN CHEMICAL SOCIETY, *List of periodicals abstracted by Chemical Abstracts*, Chemical Abstracts Service, Ohio State Univ., Columbus, 314 pp., 1956.
- AMERICAN INSTITUTE OF PHYSICS, *Style Manual*, American Institute of Physics, New York, 28 pp., 1951.
- AMERICAN MATHEMATICAL SOCIETY, A manual for authors of mathematical papers, *Bull. Am. Math. Soc.*, 49, no. 3, pt. 2, 1-16, 1943.
- EMBERGER, M. R., and M. R. HALL, *Scientific writing*, Harcourt, Brace and Co., New York, 469 pp., 1955.
- TAFT K. B., J. F. McDERMOTT, and D. O. JENSEN, *The technique of composition*, 3rd ed., Farrar and Rinehart, New York, 628 pp. 1941.
- TRELEASE, S. F., *The Scientific paper—how to prepare it, how to write it*, Williams and Wilkins Co., Baltimore, 175 pp., 1951.
- U. S. GEOLOGICAL SURVEY, *Suggestions to authors of the reports of the United States Geological Survey*, 5th ed. U. S. Govt. Printing Office, Washington, 255 pp., 1958.
- WILLIAM BYRD PRESS, *Mathematics in type*, Richmond, 58 pp., 1954.



INFORMATION CONCERNING CORPORATION MEMBERSHIP

The American Geophysical Union is a non-profit scientific organization established by the National Research Council. It is the American National Committee of the International Union of Geodesy and Geophysics, and its Executive Committee is the Committee on Geophysics of the National Research Council.

Extracts from the Statutes:

Article 3. Membership—The membership of the American Geophysical Union shall be as follows:

- (e) *Corporation Members*—Corporations and other organizations interested in geophysics elected by the Executive Committee of the Union. The designated representative of each such organization shall enjoy the privileges of a Member.

(Continued on next page)

Cut along this line

American Geophysical Union

PROPOSAL FOR CORPORATION MEMBERSHIP

To the Executive Committee, American Geophysical Union
1515 Massachusetts Ave., N.W., Washington 5, D. C.

Gentlemen:

As an indication of our interest in the aims and activities of the American Geophysical Union, and to assist in maintaining and extending its program of publication and other work in the development of the geophysical sciences, the undersigned applies for Corporation Membership in the AGU and, until further notice, agrees to pay annual dues, currently at the rate of \$100 per unit of corporation membership, in accordance with the information set forth above.

Company or Organization _____

By _____ Title _____

(Signature)

(over)

(Continued from previous page)

Extracts from the By-Laws:

- (2) . . . Members of class (e) shall pay dues of not less than \$100 for each calendar year; . . .
- (21) One copy of each issue of (a) the *Transactions*, (b) *Journal of Geophysical Research*, (c) any published *List of Members and Officers*, and (d) any other publication which may be approved for *free distribution* to the membership by the Executive Committee of the Union, shall be sent to each . . . Corporation Member. . . Each . . . organization in good standing may purchase any available publication of the Union at a discount from printed price list to non-members. The General Secretary is authorized to establish discounts for sales of publications.

Action of the Executive Committee, November 29, 1946:

- (1) A list of corporation members shall be published on one or more pages immediately after the final page of text in each issue of the *Transactions*.
- (2) A list of corporation members shall be included in the Membership Directory as a distinct unit.

AMERICAN GEOPHYSICAL UNION

1515 Massachusetts Ave., N.W.
Washington 5, D. C.

Cut along this line

Address _____

City _____

State _____

General fields of activity _____

The following person is designated as our representative in this membership _____

Title _____

Number of units of membership desired (this will be taken as one unless otherwise indicated) _____

Place _____

Date _____

Contents

(Continued from back cover)

World Maps of F_2 Critical Frequencies and Maximum Usable Frequency Factors for Use in Making Ionospheric Radio Predictions.....	Donald H. Zacharisen and Vaughn Agy	593
On the Propagation of ELF Radio Waves and the Influence of a Nonhomogeneous Ionosphere	James R. Wait	597
Night-time Equatorial Propagation at 50 Mc/s: First Results from an IGY Amateur Observing Program.....	M. P. Southworth	601
Short-Wave Fadeouts without Reported Flares.....	Howard DeMastus and Marion Wood	609
Correlated Micropulsations at Magnetic Sudden Commencements	W. K. Berthold, A. K. Harris, and H. J. Hope	613
A Proposed Experiment for the Investigation of an Energy Dependence of Photon Velocity in Vacuo	S. D. Softky and R. K. Squire	619
A New Method of Computing the Deacon Wind Profile Parameters.....	Frank L. Martin	623
Gravimetric Determination of Ocean Tide, Weddell and Ross Seas, Antarctica	Edward Thiel, A. P. Crary, Richard A. Haubrich, and John C. Behrendt	629
Simplified Method of Determining Refraction Coefficients for Sea Waves.....	R. Dorrestein	637
A New Approach to Peak Flow Estimation.....	R. Rangarajan	643
Runoff as a Function of Moisture-Storage Capacity.....	J. L. Thames and S. J. Ursic	651
Translocation of Moisture with Time in Unsaturated Soil Profiles. . .	Paul R. Nixon and G. Paul Lawless	655
Effect of Cover Types and Soils on Runoff in Northern Mississippi. . .	S. J. Ursic and J. L. Thames	663
Radioisotopes P^{32} , Be^7 , and S^{35} in the Atmosphere.....	D. Lal, Rama, and P. Z. Zutshi	669
Free Oscillations of the Earth: I. Toroidal Oscillations. . .	Freeman Gilbert and Gordon J. F. MacDonald	675
Studies in the Theory of Shock Propagation in Solids.....	William Band	695
Seismograms Associated with the Near Passage of Tornadoes.....	Carl Kisslinger	721
Study of Earthquake Mechanism by a Method of Phase Equalization Applied to Rayleigh and Love Waves.....	Keiiti Aki	729
Apparatus for Phase-Equilibrium Measurements at Pressures up to 50 Kilobars and Temperatures up to 1750°C.....	F. R. Boyd and J. L. England	741
The Quartz-Coesite Transition.....	F. R. Boyd and J. L. England	749
Elasticity of Some High-Density Crystals.....	R. K. Verma	757
Letters to the Editor:		
Changes in the Low-Rigidity Primary Cosmic Radiation During the Large Forbush Decrease of May 12, 1959.....	F. B. McDonald and W. R. Webber	767
Measurement of Radiation in the Lower Van Allen Belt	Francis E. Holly and Richard G. Johnson	771
A 'Telescope' for Soft X-Ray Astronomy.....	Riccardo Giacconi and Bruno Rossi	773
Detection of Sea-Water Motion by Nuclear Precession.....	E. L. Hahn	776
Discussion of Paper by J. D. Isherwood, 'Water-Table Recession in Tile-Drained Land'	E. G. Youngs	778
The Non-steady State of the Water Table in Drained Land.....	E. C. Childs	780
Discussion of Paper by G. W. Smith, 'The Determination of Soil Moisture under a Permanent Grass Cover'.....	Harry F. Blaney	783
Magnetic Micropulsations and the Pulsating Aurora.....	Wallace H. Campbell	784
Discussion of Paper by R. C. Bless, C. W. Gartlein, D. S. Kimball, and G. Sprague, 'Auroras, Magnetic Bays, and Protons'.....	Pierre Bernard	785
A Farside-Type Rocket Experiment for the Measurement of the Gravitational Time Effect	F. W. Lipps	786
Geomagnetic and Solar Data.....		788

Contents

INTERNATIONAL SYMPOSIUM ON ELECTRONIC DISTANCE-MEASURING TECHNIQUES

(Edited by C. A. Whitten and Erwin Schmid)

	PAGE
Introductory Remarks.....	Antonio Marussi 38
Atmospheric Limitations on Electronic Distance-Measuring Equipment <i>Moody C. Thompson, Jr., Harris B. Janes, and Frank E. Freethy</i>	38
Physical Principles of the Electro-Optical Determination of Distances.....	A. Karolus 39
The Geodimeter System: A Short Discussion of Its Principal Function and Future Development <i>Erik Bergstrand</i>	40
Use of the Geodimeter by the Coast and Geodetic Survey.....	Paul D. Thomas 41
The Model 3 Geodimeter for the Extension of Control for California Highways	James D. Carter 41
An Electro-Optical Device for Measuring Distance.....	Arne Bjerhammar 41
The 'Tellurometer' System—New Applications to Geodesy and Hydrography.....	R. D. Smith 41
Report on Electronic Distance Measurements in Australia.....	G. R. L. Rimington 43
Tellurometer Measurements in the Base Extension Network Munich.....	R. Sigl 43
Results of Tellurometer Measurements in 1958.....	Karl Gerke 44
Remarks Concerning Current Use of the Tellurometer.....	G. Coets 44
Tellurometer Trilateration in Arabia.....	Hans Meier 44
Tellurometer Operations in Topographic Mapping.....	Julius L. Speert 45
Geodetic Base Lines.....	W. A. Heiskanen 45
Measurement of Standard Base Line with the Väisälä Light-Interference Comparator <i>T. Honkasalo</i>	45
Canadian Shoran Project.....	J. E. R. Ross 46
Hiran Instrumental Developments.....	Paul W. Jordan 46
Evaluation of Hiran Networks.....	Samuel D. Owen 46
Simultaneous Adjustment of Triangulation and Trilateration: An Investigation of Tellurometer Lengths.....	L. A. Gale 47
Experimental Results of Field Tests of the Application of the Lorac Phase Comparison Radio Location System to Distance Measurement.....	B. W. Koepfel 47
Microwave Position-Fixing System.....	H. R. Smyth 49
Electronic Control Systems Used on Hydrographic Surveys.....	Gilbert R. Fish 49
Some Developments in Loran.....	W. O. Henry 50
The Method and Use of Two-Range Decca.....	G. W. LaCroix and D'A. H. Charles 51
Transistorized Raydist as Used in Geological Surveys... <i>Charles L. Drake and Walter C. Beckman</i>	52
Further Evidence of a Solar Corpuscular Influence on Large-Scale Circulation at 300 Mb <i>Norman J. Macdonald and Walter Orr Roberts</i>	52
The Ring Current and the Outer Atmosphere.....	Syun-Ichi Akasofu 53
Photographs of the High-Altitude Nuclear Explosion 'Teak'.....	W. R. Steiger and S. Matsushita 54
Balloon Observations of X-Rays in the Auroral Zone I.....	Kinsey A. Anderson 55
Some Spectral Studies of the Aurora.....	R. C. Bless, C. W. Gartlein, and G. Sprague 56
A Four-Year Summary of Whistler Activity at Washington, D. C.....	Harold E. Dinger 57
The Ion Distribution above the F_2 Maximum.....	Francis S. Johnson 57
Abnormal Features of the F_2 Region of the Ionosphere at some Southern High-Latitude Stations <i>R. G. Rastogi</i>	58

(Continued inside back cover)